COMMONWEALTH OF MASSACHUSETTS
EXECUTIVE OFFICE OF ENVIRONMENTAL AFFAIRS
DEPARTMENT OF ENVIRONMENTAL PROTECTION
ONE WINTER STREET, BOSTON, MA 02108 617-292-5500

MITT ROMNEY
Governor

KERRY HEALEY
Lieutenant Governor

ELLEN ROY HERZFELDER
Secretary

ROBERT W. GOLLEDGE, Jr.
Commissioner

July 21, 2003

David Conroy
U.S. EPA – New England
One Congress Street
Boston, MA

Dear Mr. Conroy:

Enclosed please find "The MA31 Conversion Factor Analysis and Interim Test Effectiveness Evaluation." This report was required in your January 16, 2001, approval of our State Implementation Plan (SIP) for the Massachusetts Enhanced Inspection and Maintenance Program (65FR 69254 et seq.) This report is based on a study conducted by Sierra Research, Inc., and its subcontractor, Gordon-Darby, Inc.

Our implementation of this study encountered a number of delays, including equipment availability and set up time, and the accommodation of the needs of our data collection effort and those of the on-going inspection program in Arizona. During this period, we discussed progress on the evaluation with members of your staff as we completed various phases of the field work and analysis. The data that forms the basis for this report will also be used in our evaluation of how well the Massachusetts program is working in the field, an evaluation that EPA also requires.

This study was performed to fulfill EPA's requirement that decentralized Inspection and Maintenance programs using alternative testing equipment measure how effectively their test identifies excessively polluting vehicles, compared to EPA's benchmark IM240 test (40 CFR 52). When Massachusetts designed its I&M program in the mid 1990s, we decided not to adopt the EPA IM240 test. Instead, we proposed a test that has less expensive analytical equipment and shorter testing times than IM240. These choices helped keep the Massachusetts inspection fee that motorists pay below what most other states charge, cut equipment costs to inspection stations almost in half, and reduced the amount of time that motorists spend traveling for and getting inspections. These choices caused the Massachusetts test to be somewhat less effective than the federal benchmark IM240 test, while still meeting Massachusetts' air quality goals.

This study examined the Massachusetts test (“MA31”) under controlled conditions, and found that the test that Massachusetts motorists take today exceeds EPA’s test effectiveness targets for 2 out of 3 pollutants (hydrocarbons and carbon monoxide), and achieves about 9/10ths of the test effectiveness target for the third (NOx.) The study recommended that:

- DEP adjust the calibration of its emissions test results to more closely match EPA's standard measurement of grams per mile. One adjustment, which was made in July 2001, made the Massachusetts test more accurate for hydrocarbons, Carbon Monoxide, and NOx compared to EPA's benchmark IM240 test. The study also recommends that a second adjustment be made to further improve the test’s accuracy. DEP is now reviewing this recommendation and will be consulting with your staff prior to finalizing our decision, expected by October 1, 2003.

- The Massachusetts test needs to more effectively identify NOx emissions, and the study contained recommendations that would accomplish this goal.

We look forward to discussing the study results and recommendations, and the upcoming program evaluation with you, and to working with our Inspection and Maintenance Advisory Committee on the continued improvement of the Enhanced Emission Inspection Program.

If you have any questions or need additional information about this study, please feel free to contact me at 617-556-1020.

Sincerely,

Nancy L. Seidman, Director
Division of Consumer and Transportation Programs
Bureau of Waste Prevention

cc:
James C. Colman, BWP

SUMMARY

OF THE

MA31 CONVERSION FACTOR ANALYSIS AND

INTERIM TEST EFFECTIVENESS EVALUATION

JULY 21, 2003

ABSTRACT

On November 16, 2000, EPA gave final approval to the current Massachusetts I&M Program. Approval of the I&M State Implementation Plan recognized that the Massachusetts test equipment was less effective than EPA's benchmark test (IM240), established interim "test effectiveness" targets that the Massachusetts test was required to meet, required that Massachusetts conduct a study to verify whether those targets were being reached, and stated that EPA would take action to make up for any shortfall if those targets were not reached.

That study – "MA31 Conversion Factor Analysis and Interim Test Effectiveness Evaluation" - is being submitted to EPA today to fulfill that EPA requirement. In order to ensure that the study met EPA's expectations, EPA and DEP cooperatively designed the study.

The study had three objectives:

1. Evaluate the conversion factors used by the Massachusetts program to express test results in terms that are comparable to IM240, and recommend improvements if needed so that Massachusetts test results more closely approximate results that would be expected from the IM240 test.

2. Evaluate the effectiveness of the MA31 drive trace relative to the IM240 drive trace. "Test Effectiveness" measures how well a state's I&M test equipment and drive trace identify excess emissions from a vehicle fleet compared to EPA's benchmark equipment and drive trace, IM240.

3. Evaluate potential program changes that could increase the MA31 test effectiveness, if needed.

The study found that:

1. The conversion factors needed to be revised. In **July 2001**, DEP revised the conversion factors to make the test more accurate. That change also had the effect of decreasing failure rates. A final change to the conversion factors will be implemented by DEP after consultation with EPA and discussing the proposed changes with stakeholders. This change, which DEP expects to make by October 2003, will also increase NOx effectiveness.

2. The effectiveness of the Massachusetts test after the change in conversion factors in July 2001 exceeded target levels for HC and CO, and was below the target level for NOx. In **March 2003**, DEP made a program change that increased test effectiveness as follows:
    - HC: 91% compared to 85% target
    - CO: 93% compared to 87% target
    - NOx: 75% compared to 85% target

3. DEP should improve test effectiveness for NOx, and should consider implementing a longer test or changing the pass/fail cutpoints to make the Massachusetts test stricter.

DETERMINATION OF POLLUTANT-SPECIFIC CONVERSION FACTORS
FOR
MASSACHUSETTS-TO-IM240 CONVERSION OF EMISSIONS TEST RESULTS

INTRODUCTION

Why Does Massachusetts Have an I&M Program?

Massachusetts continues to violate minimum federal standards for ground-level ozone pollution. On our "bad air" days there is an increase in asthma attacks and hospitalizations for severe respiratory ailments. In order to reduce the number of "bad air" days and to comply with the US Clean Air Act and EPA regulations, Massachusetts must implement various federally mandated programs. To reduce pollution from motor vehicles, Massachusetts is required to operate an Enhanced Inspection and Maintenance (I&M) program. I&M programs identify onroad gasoline motor vehicles that substantially exceed their designed emissions levels by periodically testing them. Repairs are required for vehicles that fail the emissions test. The U.S. Environmental Protection Agency (EPA) sets minimum standards for I&M programs[1].

States include their I&M programs in a State Implementation Plan (SIP), a federally-enforceable binding commitment from the state to EPA, that lists the programs the state will use to meet federal pollution reduction goals. These programs include reductions from many sources, such as power plants, small businesses, factories, construction equipment and onroad vehicles. For I&M programs, EPA establishes different levels of credit to states for reducing pollution based on the kind of I&M test and program the state implements.

How Does an I&M Program Work?

Recent – or "Enhanced" – I&M programs require placing vehicles on a dynamometer, a treadmill-like device that puts resistance against the tires to simulate onroad driving. Then vehicles are accelerated and decelerated according to a prescribed driving pattern ("drive trace"), and tailpipe pollutant levels are collected and recorded. Pollutant readings for hydrocarbons (HC), Carbon Monoxide (CO) and Oxides of Nitrogen (NOx) are compared against each pollutant's pass/fail points.

[1] 40 CFR Part 51, Subpart S (§51.350 et seq.).

What is EPA's Model I&M Program?

EPA developed a "model" enhanced I&M test, called IM240. This model program tests vehicles using laboratory-grade instruments in stand-alone contractor-run buildings. In these centers, which only perform emissions testing (i.e., no safety inspections or repairs), emissions are continually measured for 4 minutes on a dynamometer. The IM240 dynamometer drive trace is derived from the EPA test used to certify new vehicles, and covers almost 2 miles of simulated driving at speeds of up to 57 mph.

Many states believed that this IM240 model was too expensive, was inconvenient for motorists, or did not fit well with the existing businesses that performed emissions tests. In some states, early implementation of the model IM240 program was a failure: long lines, motorist revolts and the canceling of multi-million dollar contracts. In response to this situation, Congress in 1995 passed a law that required EPA to grant states flexibility to select alternative tests and programs.

EPA's IM240 guidance proposes two different pass/fail points (cutpoints). EPA allows states to choose either the less stringent "start-up" cutpoints or the more stringent "final". Both sets of cutpoints are designed to identify vehicles with malfunctioning emission control systems that pollute much more than they were designed to. EPA final cutpoints are stricter and obtain more pollution reductions. EPA start-up cutpoints are more lenient and obtain fewer reductions. Massachusetts uses the more lenient cutpoints because its air quality plans (the SIP) did not require greater air pollution reductions from motorists.

How are alternative I&M programs different than EPA's IM240 model?

One difference in the programs is where the testing is located. EPA envisioned the creation of a relatively small number of stand-alone emissions test centers and requiring all motorists to get tested there. That approval would have taken the emissions testing business out of local, conveniently located inspection shops and forced motorists to drive to one of only 60 locations statewide for testing. Additionally, the question of how to accommodate safety inspections with an IM240 emissions program was complicated. Safety inspections would either be taken away from existing businesses or motorists would have to drive to two different locations for the tests.

Another difference is the effectiveness of alternative tests. EPA considers these alternative programs to be less effective than the IM240 model because the drive traces tend to be shorter and at lower speeds, the testing equipment is usually cheaper and less accurate, and because it believes that tests are delivered more accurately in test-only businesses compared to businesses that also repair vehicles. Because EPA considers these alternative tests to be less effective than IM240, EPA assigns less emissions credit to states with alternative I&M tests.

Why Didn't Massachusetts used the EPA Model Test, IM240?

The Department of Environmental Protection (DEP) and the Registry of Motor Vehicles (RMV) are jointly responsible for implementation of the Massachusetts Enhanced Emission and Safety Test. Their goals are to provide a comprehensive test that is convenient to motorists, works well in local inspection shops, provides the emission reductions needed for the SIP, and ensures vehicle safety.

In the mid-1990s, after Maine, Pennsylvania, and Texas had abandoned their centralized IM240 programs, the Administration directed the DEP and the RMV to examine the possibility of implementing an alternative test in Massachusetts.

After conducting an "I&M Summit" with over 30 state officials from 11 different state agencies, and holding a workshop with I&M business and equipment officials, national trade associations, and other stakeholders, the DEP and RMV decided that the Massachusetts I&M program would have to balance as well as possible the achievement of three different goals:

- It had to be convenient in terms of price and location for Massachusetts motorists.
- It had to fit well with the business plans of the inspection and repair industries, ensuring that the private sector could still deliver this program.
- It had to achieve significant pollution reductions.

Development of an alternative test and a compatible contract model took over 2 years. It included technical research by DEP and RMV staff, thoughtful review and consideration by a large and diverse I&M Advisory Committee, publication of a report highlighting the kind of choices that Massachusetts was considering, and passage of legislation in late 1997 that permitted DEP and RMV to proceed.

Implementing IM240 in Massachusetts would have cost inspection stations significantly more to buy or lease equipment ($55,000 to $65,000 per test lane compared to the current $36,000). If the IM240 test were kept in local shops, the inspection fee would have cost up to $50; it is currently $29. If the IM240 test were instead conducted in stand-alone test centers, motorists would have had to drive much greater distances for the test. Safety inspections would either be taken from local business and placed in the test centers, or motorists would have had to drive to different locations to obtain the two tests.

It takes 4 minutes to measure emissions on the IM240 test at speeds up to 57 mph. It takes from 1 1/2 to 3 minutes to measure emissions on the MA 31 test at speeds limited to 30 miles per hour. While the IM240 test would have provided a higher level of accuracy and emission reductions, it did not meet the goals that stakeholders had developed for Massachusetts.

What Advantages Did the Alternative Test Offer Massachusetts?

The use of an alternative test helped optimize customer convenience in two ways. First, it offered a faster test that would reduce wait times for motorists. The difference in test times between the MA31 test and the EPA IM240 as actually delivered in the field – including set up and measuring emissions - could range from 3 ½ to 5 minutes. While that is a small increment for a single vehicle, the cumulative effect on wait times for a line of cars can be significant for people at the end of the line

Second, the less expensive equipment helped keep the motorists' fee below many other states. With many states I&M fees ranging from $40 to $50, the cumulative savings over our $29 fee for the estimated 4.2 million vehicles is significant.

The use of an alternative test also fit better with the business plans of the Massachusetts inspection industry. Operating the dynamometers at lower speeds (31 miles per hour vs. the IM240's top speed of 57 miles per hour) addressed industry concerns about noise and safety in a small shop environment. The cost of equipment was of great concern to the inspection industry. In addition, every minute saved on an emissions test was worth about $1 to the inspection stations. At 2-3 emissions tests per hour, stations viewed the savings as significant.

EPA – which approved the use of the Massachusetts, New York and Rhode Island alternative tests – believed that the equipment and trace were capable of achieving the level of pollution reductions described in Massachusetts' pollution reduction plans. Similar equipment is currently being used in New York and Rhode Island.

What Alternative Test Did Massachusetts Develop?

To maximize customer convenience, DEP and the RMV decided to keep emissions and safety testing together, and to keep the combined test in local inspection stations, convenient to where people live and work. There are about 1600 public testing stations available now, instead of the 60 or so we would have had under EPA's plan.

To establish the pass/fail points for the program, Massachusetts decided to use EPA's less-stringent IM240 start-up cutpoints. Although these less-stringent cutpoints would result in fewer high-emitting vehicles being failed than the IM240 test, the emission reductions from the I&M program would still be sufficient to meet Massachusetts' SIP requirements.[2]

DEP and the RMV worked through an active I&M Advisory Committee composed of inspectors, educators, environmentalists, EPA and repairers to identify and develop a program and a test that best met the Commonwealth's goals and that the stakeholders believed would work well in Massachusetts.

The equipment that Massachusetts chose for its alternative test had less costly and less accurate pollution measuring equipment than the more costly laboratory grade IM240 equipment. This equipment, known as MASS99, was the same equipment selected for New York's I&M test, and is currently used in NY and RI.

The drive trace that Massachusetts selected is a shorter, lower-speed 31 second dynamometer drive trace (MA31), with a top speed of 30 mph. The trace is the same as the "BAR31" alternative drive trace developed by California's Bureau of Automotive Repair (BAR), and is also used by Oregon and Rhode Island.

These two test features – less expensive repair-grade equipment and a shorter, lower speed drive trace - are together called the "MA31 test".

[2] States typically use much looser cutpoints during program implementation to accommodate any bugs in the testing system or in the I&M program. As the program is implemented, states typically change their cutpoints in a series of steps until the program's final cutpoints are reached.

How Does the I&M Test Work?

When a vehicle arrives for an enhanced I&M test, the inspector enters vehicle information into the test's computer system. Two key test features hinge on proper vehicle identification:

1. Dynamometer loading – for an accurate test, it is necessary for the dynamometer to place the proper resistance against vehicle's tires. The dynamometer loading calculations come from the EPA test used to originally certify that the vehicle meets "new car" emission standards. They vary by type and weight of vehicle.

2. Cutpoints (pass/fail points) – to determine whether a vehicle has passed or failed, it is necessary that test results be compared to the correct cutpoint. Cutpoints vary by vehicle type [car vs. truck], model year, and, for trucks, by weight category.

Once the vehicle is ready for testing, an inspector "drives" the vehicle on the dynamometer while the emissions and exhaust flow are collected and analyzed. At the conclusion of the test, the test system corrects for any differences from the IM240 equipment and drive trace, calculates the emissions and the distance driven on the dynamometer, and reports emissions in EPA IM240 standard grams per mile. The grams per mile results are then compared to the cutpoints (which are also expressed in EPA IM240 grams per mile), and a pass/fail determination is made.

How Do Results From MA31 and IM240 Compare?

When a vehicle is tested using IM240 test equipment and the IM240 dynamometer drive trace (trace), emissions for each pollutant are measured and reported in EPA IM240 standard grams per mile for each pollutant. The emissions are compared to IM240 cutpoints to make a pass/fail determination.

When a vehicle is tested using the MA31 test (using MASS99 test equipment and the MA31 trace), emissions are first measured as a concentration (parts per million) and are then converted to a mass measurement (grams per mile). Then the software converts the amounts for each pollutant into EPA IM240 standard grams per mile before being compared to IM240 start-up cutpoints for a pass/fail determination. The initial reporting of emissions as a concentration and subsequent conversion to grams per mile is a byproduct of using the less expensive analytical equipment used in Massachusetts, Rhode Island and New York.

The software conversion of "raw" grams per mile readings into EPA IM240 standard grams per mile readings is necessary for two reasons. First, the MASS99 equipment has different measurement characteristics than IM240 equipment. Second, vehicles tested using the MA 31 drive trace are being driven on a different "road" than the IM240 drive trace. Just as driving on a different road will yield different "miles per gallon" figures, using a different drive trace yields different emissions readings. Since the Massachusetts pass/fail points are expressed in EPA IM240 standard grams per mile, the readings for this alternative drive trace must be converted to EPA IM240 standard grams per mile.

How Are MA31 Test Results Converted to IM240 Standard Grams per Mile?

Almost all states that use an alternative I&M test convert their emission test results to EPA IM240 standard grams per mile for each pollutant. The vehicles emissions are then compared to the EPA cutpoints (also expressed in EPA IM240 standard grams per mile), and a pass/fail determination is made. The conversion can account for either a different drive trace, different equipment or both. Massachusetts uses both a different trace and different equipment than IM240.

Some states choose not to express emissions readings in EPA IM240 grams per mile. Rhode Island, for example, instead changes its cutpoints to account for its differences in drive trace and equipment.

The only software-based conversion factors available from other states at the beginning of the Massachusetts I&M program corrected for different equipment. DEP used equipment correction factors from the New York Department of Environmental Conservation, which uses the same equipment as Massachusetts.

Since there were no software-based conversion factors available to account for the differences in the MA31 drive trace, Massachusetts made three revisions: 1) adjusted its start-up cutpoints to accommodate the different trace; 2) added the NY correction factors to accommodate the different equipment; and 3) planned to conduct a study in the future that would determine a combined software correction factors for both the trace and the equipment.

Why Change the Conversion Factors?

EPA conditionally approved the use of the conversion factors at the start-up of the Massachusetts I&M program, and stipulated that DEP must develop Massachusetts-specific conversion factors. In response to that requirement, DEP and EPA agreed that Massachusetts would conduct a study with two purposes: Massachusetts would be able to increase the accuracy of its test, and EPA would obtain information that it could use to establish final test effectiveness numbers for the Massachusetts test. In order to ensure that both needs were met, DEP sought EPA's assistance during the development of the study. EPA reviewed and accepted the study design.

CONVERSION FACTOR STUDY

In response to EPA's requirement, and to ensure that there was one software-based conversion factor for each pollutant that accommodated differences in both test equipment and drive trace, DEP initiated a study to compare the emission results from the IM240 and MA31 tests. The study and many of the methodological technical details were designed in cooperation with EPA staff. The results were designed to provide one conversion factor for each pollutant that would convert emissions gathered from the MA31 test into EPA IM240 standard grams per mile.

How Was the Study Conducted?

In October 2000 DEP contracted with Sierra Research of Sacramento, California (Sierra) to perform the study. Sierra subcontracted with Gordon-Darby, Inc of Louisville, KY (Gordon-Darby) to test vehicles in Arizona using the IM240 and MA31 tests. Gordon-Darby operates the emissions testing program for the greater Phoenix, Arizona area, widely considered to best represent IM240. Agbar Technologies, the contractor administering the Massachusetts I&M program, was responsible for providing and maintaining a MASS99 test equipment for the study.

Vehicles were tested on the IM240 equipment first using an IM240 trace. This was followed by a test on the IM240 equipment using the MA31 trace. Vehicles were then moved to the MASS99 test equipment where they were tested using the MA31 trace, followed by an IM240 trace. This gave each test vehicle every possible combination of test equipment and trace.

How Were Vehicles Selected for the Study?

Vehicles that showed up for their routine Arizona emissions inspection were candidates for study participation. To provide a diverse enough set of data to evaluate the conversion factors, it was necessary to test a variety of vehicle types and model years with various emission rates, including older dirtier vehicles with high emissions. The study design, which had been reviewed by EPA, established the goal of recruiting equal numbers of vehicles across model years, vehicle types, and level of emissions.

Testing for the study was performed from December 2000 to August 2001. There were 612 vehicles tested that met selection criteria and yielded valid tests.

CONVERSION FACTOR ANALYSIS

How Were the New Conversion Factors Calculated?

The emission results from the IM240 and MA31 tests are compared using a linear regression. This type of analysis plots the emission results and then forms a line that best fits the data points so that all the points are as close to the line as possible. The mathematical slope of the line becomes the conversion factor.

This type of analysis is necessary to match the MASS99 equipment's emission calculations, which are designed for a linear conversion of emission scores. The linear equation used in the MASS99 system is:

IM240 Equivalent Emissions = Raw MA31 Emissions X Conversion Factor

Since the main purpose of the conversion factors is to make accurate pass/fail determinations, it is critical that the linear regressions generating the conversion factors for each pollutant are as accurate as possible when emissions are near the cutpoints.

Test accuracy is much more important for vehicles that pollute at levels near the cutpoints. For those vehicles, normal variations in accuracy for the test coupled with normal variations in the amount of pollution emitted by any one vehicle can be enough to cause the vehicle to pass a test once, then fail a second time.

In contrast, normal variances in test accuracy for clean vehicles - for example, those vehicles that emit 10 times below the cutpoints - is much less important because those clean cars will still pass. Likewise, normal variances in test accuracy from gross polluters – for example, those vehicles that pollute 10 times above the cutpoints - is also much less important because those extremely dirty vehicles will still fail.

To ensure that the correction factor analysis was not biased by extremely high emitters, the study excluded data from the extremely high emitting vehicles for some pollutants.

Why Were Interim Conversion Factors Calculated?

At the beginning of the program, DEP installed conversion factors in the software that only corrected for differences in the equipment, and adjusted the start-up cutpoints to correct for differences in the trace. Since DEP's final cutpoints did not include any correction for the drive trace, and since DEP moved to its final cutpoints in April 2001, DEP decided to analyze a preliminary dataset of 341 vehicles.

What Were the Outcomes of the Interim Conversion Factor Analysis?

Table 1 compares the interim conversion factors to those initially used in the software. As can be seen, there was a substantial disparity between the two sets of conversion factors. The disparity was causing the reported emissions from vehicles *to be higher* than EPA IM240 standard grams per mile, which would cause motorists' emission readings to be inflated. This overestimation would also cause motorists to fail when they should have passed.

Following this analysis, on July 11, 2001, the conversion factors were changed to the interim values calculated from the preliminary 341-vehicle dataset. The change was made to increase the accuracy of the test and reduce false failures. It also had the effect of lowering the failure rates

TABLE 1
COMPARISON OF INITIAL AND INTERIM CONVERSION FACTORS

	HC	CO	NOx
Massachusetts I&M Program – Initial	1.50	0.86	0.86
Interim – based on AZ 341 Sample Dataset	0.98	0.57	0.56

There were three outcomes from this change:

1. MA31 test accuracy was improved;
2. emissions readings for vehicles were lower after the change; and
3. the emissions failure rate declined from a range of 7%-9% to 4% - 6%.

Why Were the Conversion Factors So Different?

The initial software conversion factors accommodated only differences in the equipment. The drive trace differences were accommodated by adjustments to the initial cutpoints. As of April 2001, however, with the implementation of the final cutpoints, the cutpoints no longer included any correction for the drive trace. The interim conversion calculations accommodate both differences in equipment and differences in trace.

Final Conversion Factors

Following completion of this study in August 2001, conversion factors for each pollutant were calculated for the 612 vehicle sample dataset. While the original study called for 1000 vehicles to be tested, a variety of factors, which are discussed in the technical report, led to a smaller sample size. That reduction was discussed and accepted by EPA. To improve the accuracy of emission calculations for vehicles emitting near the cutpoints, the analysis of the complete data set excluded vehicles with raw MA31 emissions more than 1.5 times each pollutant's maximum cutpoint, as was done with the interim conversion factors. Emissions were high enough to fail both tests.

Table 2 compares conversion factors developed from both the 612 and 341 vehicle datasets to the initial factors used in the program.

TABLE 2
COMPARISON OF CONVERSION FACTORS

	HC	CO	NOx
Massachusetts I&M Program Initial	1.50	0.86	0.86
Interim, Implemented July 2001 (based on AZ 341 Sample Dataset)	0.98	0.57	0.56
Final (based on AZ 612 Sample Dataset)	0.87	0.53	0.60

As Table 2 shows, the final conversion factors developed from the 612 vehicle dataset are close to the interim conversion factors. The final conversion factors for HC and CO are slightly lower, meaning the interim conversion factors still cause emissions results reported by MA31 to be slightly higher than EPA IM240 standard grams per mile, although they represent a substantial improvement to test accuracy compared to the initial conversion factors.

For NOx, the final conversion factor is slightly higher than the interim conversion factor, meaning the interim conversion factor results in NOx emissions being slightly lower than EPA IM240 standard grams per mile. Similar to HC and CO, the interim NOx conversion factor still represents a substantial improvement in test accuracy compared to the initial conversion factor.

When Will the Final Conversion Factors Be Implemented?

The final conversion factors will be implemented after final quality assurance of the data and the conversion calculations is complete and stakeholders are provided advance notice of these changes. Changing the conversion factors also affects other components of the program, such as the recently implemented dilution check.

In addition, other software upgrades have been in progress. Attempting too many simultaneous changes constitutes poor network management, increasing the potential for network disruption. Once the final conversion factors are verified and stakeholders consulted, DEP will implement the changes at a time that minimizes potential for network disruption.

INTERIM TEST EFFECTIVENESS EVALUATION

<u>What Is Test Effectiveness?</u>

"Test effectiveness" is a measure of how well a state's I&M test equipment and drive trace identify excess emissions from all the states' vehicles compared to EPA's "gold standard" test equipment and drive trace, the IM240.

Test effectiveness numbers are derived from studies that compare the state's equipment and drive trace under ideal, controlled conditions to the IM240 equipment and drive trace under ideal, controlled conditions.

These studies then calculate a test effectiveness percentage for each of the three tested pollutants: hydrocarbons (HC), oxides of Nitrogen (NOx), and carbon monoxide (CO). "100% test effectiveness" would match the test effectiveness of EPA's standard, the IM240 test and equipment.

This test effectiveness comparison is deliberately done under ideal conditions in order to separately analyze how well the equipment and the trace perform. In contrast to "test effectiveness", "program effectiveness" examines how well the equipment and trace perform as they are used in the field, and is usually analyzed as a part of a full program evaluation[3].

"Excess emissions" is the difference between how much a failed vehicle pollutes when tested and how much it pollutes after repairs are done. Typically, repairs on a failed vehicle will reduce emissions well below the cutpoints. That is because most vehicles in good repair will pass the test by a large margin, and most vehicles with malfunctioning emissions control systems will fail the test by a large margin. Repairs made before the test occurs are not counted as excess emissions, because the test did not identify them. In that case, the air is cleaner, but the I&M program does not get credit for that reduction.

When analyzing test effectiveness – how well the equipment and drive trace perform in controlled circumstances – EPA assumes vehicle pollution is reduced only to the cutpoints. When analyzing program effectiveness – how well the program works in the field – EPA analyzes data that show the actual reductions achieved from repairs.

Test effectiveness numbers are used by EPA and states as one of the many inputs in EPA's computerized models for estimating air pollution ("models"). There are dozens of different inputs that describe, for example, ambient monthly temperatures, average vehicle miles traveled, average age and profile of the vehicles in the state, and so on. These "models" are used to create a system of air pollution "credits" linked to reductions of pollution from air pollution sources (e.g., vehicles, power plants, other businesses). EPA and states then use these "models" to establish state pollution reduction targets, and to develop estimates of "credits" that come from each I&M program.

---

[3] A program evaluation analyzes how well the entire program works in the field, including factors like in-use accuracy of the MASS99 equipment and software, amount of permitted variance in the driving trace, inspector's ability to match the driving trace, amount of pre-test repairs, the effect of improper test administration by the inspector, the effect of motorist non-compliance, et cetera.

EPA periodically changes its models to reflect greater knowledge about how vehicles pollute. For example, EPA's MOBILE 6 model, released in 2001, revised and lowered estimates of the benefit from I&M programs to reflect the fact that the emissions control systems of vehicles manufactured in the 1990s tended to perform better than those manufactured in the 1980s (data from 1980s vehicles was the basis for EPA's MOBILE 5 model). In other words, after the Massachusetts I&M program started, EPA lowered its estimate of the effectiveness of the program because vehicles were in general running cleaner. While the I&M program stayed the same, EPA's new model gives it less credit.

What Test Effectiveness Percentages Did EPA initially assign to the MA 31 Test?

When EPA approved the Massachusetts I&M program, it included the assignment of test effectiveness percentages as follows:

- HC (hydrocarbons): target is 85% as effective as IM240
- CO (carbon monoxide): target is 87% as effective as IM240
- NOx (oxides of nitrogen): target is 85% as effective as IM240

In other words, EPA predicted that the shorter MA31 drive trace and less expensive equipment would, under ideal conditions, be less effective at identifying vehicles with broken emission control systems than EPA's gold standard IM240 trace and equipment.

EPA's initial calculation was based in part on New York data and in part on professional judgment. Since the New York data was from a small number of vehicles, and since the NY I&M test used the same equipment but a different drive trace than Massachusetts, EPA required Massachusetts to conduct a test effectiveness study to develop more accurate test effectiveness percentages for MA 31.

How Was Test Effectiveness Analyzed?

DEP estimated excess emissions identified by the MA31 test and the IM240 test, compared those results, used that comparison to estimate how well MA31 performed in comparison to IM 240. Then DEP compared the resulting test effectiveness percentages for the MA31 test to Massachusetts' targets. The correlation factor study described earlier in this summary was also used as part of this Test Effectiveness work.

In order to do this, DEP obtained vehicle data from an actual working I&M program that used the IM240 test, and applied appropriate statistical methods to provide results more representative of Massachusetts vehicles, as follows:

- A random sample of IM240 tests from Arizona's ongoing I&M program was selected;
- The IM240 records from the Arizona vehicles were obtained;

- MA31 results for those vehicles were estimated based upon calculations derived from the 612 car dataset;

- a statistical method ("Monte Carlo") was used to simulate the varied kinds of test results found in operating I&M programs; and

- the excess emissions identified in the IM240 and MA 31 datasets were compared.

DEP believes that this analysis is sound and that it is sufficient for its intended purpose: to provide an estimate of MA31 test effectiveness and to suggest ways to improve effectiveness. Details about this methodology are contained in the "MA31 Conversion Factor Analysis and Interim Test Effectiveness Evaluation" (the Technical Report). On July 21, 2003, DEP submitted the Technical Report to EPA and requested that EPA review DEP's work and provide suggestions to improve test effectiveness. DEP will also discuss the Technical Report with I&M program stakeholders.

What Are the Major Limitations of the Study?

All studies have limitations. Both the 2%AZ sample and the 612 vehicle dataset varied somewhat from the actual Massachusetts fleet. For example, the distribution of vehicles in the Arizona fleet contained more older vehicles than are tested in Massachusetts. In addition, the 612 vehicle data set study was designed to obtain relatively equal numbers of vehicles across model years, vehicle types and rates of emissions. It was not designed to be representative of the actual distribution of Massachusetts vehicles by model year.

Appropriate statistical methods were used in an effort to make the results more representative. Since EPA's original selection of numbers for test effectiveness was an estimate, this study's results very likely yield more accurate test effectiveness figures.

A second limitation –differences in results seen when vehicles are tested multiple times on the same equipment – affects all I&M programs. This "test-to-test" variance comes from the vehicle being tested (vehicle variance) and from inherent limitations of the test.

A single motor vehicle tested multiple times, driven precisely the same way each time, on equipment that functioned perfectly each time, would yield varying pollution readings. Those differences appear to be caused by the vehicle itself producing different amounts of pollution, even when driven in exactly the same manner.

The test itself also introduces additional test to test variations. There is always some difference in how the driving trace is conducted; precise duplication is impossible with human drivers. In addition, the equipment that analyzes the emission gases has margins of error. For example, the less expensive chemi-luminescense equipment used to measure NOx in Massachusetts has a greater margin of error than the Flame Ionization Detector used in the more expensive IM240 equipment.

Add these variances together, and it is normal for I&M test results to vary for the same vehicle tested on the same equipment.

To offer an analogy, it’s similar to taking one’s pulse. If someone were to sit still in a chair and have their heart rate counted for one minute, they would get a specific pulse rate. Conducting that test many times would yield different results, since hearts sometimes beat slower and sometimes faster. Each result is accurate at the time it is taken; but each yields different results because the heart’s pulse rate varies.

Now, instead of counting the pulse for one full minute, what if someone counted the number of beats for six seconds and multiplied the result by ten to get a pulse rate (in other words, use a shorter, faster, and more convenient test). This quicker test introduces even more variability into the measurement of an accurate pulse rate.

That kind of faster, more convenient pulse taking is common, because it counts well enough to identify major pulse rate problems (for example, an extremely fast heartbeat).

I&M tests only need to be good enough to identify major pollution problems: the dirty, broken vehicles that are polluting much more than they were designed to.

If a clean car, polluting a 1/10th of the pass/fail points, has varied results from test to test, the vehicle is still clean and still passes. If a grossly polluting vehicle, emitting at 10 times the pass/fail points, has varied results from test to test, the vehicle is still dirty and still fails.

However, a vehicle that pollutes right at the pass/fail points is a different matter. That vehicle has a malfunctioning emission control system, and its own “vehicle variance” alone will cause it to sometimes pass and sometimes fail. Add in the variance from the test itself, and it’s expected that some vehicles will be clean enough to pass one time, then dirty enough to fail the next.

This “vehicle variance” is common in every I&M program across the country, including EPA’s benchmark IM240 test. That factor alone can cause the vehicles that pollute right at the cutpoints to fail an I&M test one day and pass the next.

When developing and implementing I&M programs, states balance the added cost to motorists and inspection stations of the laboratory grade equipment that has some variance with less expensive equipment that has somewhat more variance. Some states chose the relatively more accurate and more expensive equipment. New York, Rhode Island, Massachusetts and other states chose less expensive equipment.

What Were the Results of the Interim Test Effectiveness Analysis?

The test effectiveness numbers predicted for the MA31 test and equipment after the correction of conversion factors in **July 2001** were:

- HC: 87% as effective as IM240 compared to the 85% target
- CO: 90% as effective as IM240 compared to the 87% target
- NOx: 69% as effective as IM240 compared to the 85% target.

In other words, the Massachusetts test was predicted to exceed its test effectiveness targets for HC and CO and to obtain about 8/10ths of the target for NOx.

What Were The Interim Report's Recommendations?

For the short term, the recommendation was to delete "Fast Pass" from the MA31 test. "Fast Pass" was implemented to reduce the amount of test time for motorists and inspectors. It allowed a vehicle to pass if its emissions results were below the pass/fail points on any one of six times the 31 second test was given. In other words, if a vehicle with "true" emissions results above the pass/fail points dipped below those points on any one of up to six chances to pass – due to vehicle variance - the vehicle was passed. This lowered test effectiveness because it tended to allow more polluting cars to pass.

The analysis predicted that substituting "Fast Pass" (six chances to pass) with a more protective test (two chances to pass) would produce these test effectiveness results:

- HC: 91% as effective as IM240 compared to the 85% target
- CO: 93% as effective as IM240 compared to the 87% target
- NOx: 75% as effective as IM240 compared to the 85% target

In other words, by eliminating "Fast Pass" the Massachusetts test was predicted to exceed its targets for 2 out of 3 pollutants, and to obtain about 9/10ths of the target for NOx. This change was made in **March 2003**. Implementing the final change to the conversion factors will further improve test effectiveness for NOx and slightly reduce test effectiveness for HC and CO.

For the long term, the recommendation is to evaluate and select methods that will increase NOx effectiveness. These actions will increase the program's failure rate. Possibilities include:

- Switch from the short MA31 test to MA147 (a longer, higher speed test) for all vehicles that are tested on the dynamometer
- Use a hybrid: the shorter MA31, with a longer MA147 as a "second chance to pass"
- Lower pass/fail points for the MA31 test (requires a regulatory change)

What Actions Were Taken?

In **March 2003** DEP deleted fast pass in order to increase predicted test effectiveness for NOx from 69% of IM240 to 75% of IM240.

What Are the Next Steps to Improve Test Effectiveness?

DEP will discuss this Interim Effectiveness Report with EPA and other stakeholders to obtain their input and recommendations.

DEP will propose, discuss and implement improvements to the test designed to raise test effectiveness for NOx to the 85% target. While the schedule will be dependent upon the method selected, DEP intends to select a method in summer 2003 and implement the improvement before Spring 2004.

Do the NOx Test Effectiveness Findings Have an Impact on Massachusetts' Ability to Meet its Clean Air Goals?

Massachusetts does not attain the federal standards for ozone, and has established (and obtained EPA approval) of a State Implementation Plan (SIP) that commits to implementation of strategies for reducing ozone pollution and attaining the federal standards. NOx is a significant contributor to ozone pollution.

The major elements of Massachusetts' strategy for reducing NOx pollution are the summertime NOx cap and trade program for power plants (which provides about half of the needed reductions), the "Low Emission Vehicle" Program in which new vehicles purchased in Massachusetts are required to meet California's standards for clean engines (which provides about a quarter of the needed reductions), and the Enhanced Emission Test Program (which is expected to provide the remaining quarter of needed NOx reductions).

This study demonstrates that Massachusetts equipment and drive trace currently obtains about 90% of the test effectiveness for NOx that is expected from the IM test. The study also identified several possible modifications of the Massachusetts emissions test that would bring the test closer to the test effectiveness targets established by EPA in its approval of the Massachusetts' IM SIP.

Massachusetts will discuss with EPA the report findings, what effect the current NOx test effectiveness percentages may have on our pollution reduction plans, and what actions may be needed to meet the SIP requirements.

I&M GLOSSARY

Measuring Emissions

"Conversion Factors" – Adjustments made to convert raw emissions results from an emissions test to EPA IM240 standard gpm. Conversion factors compensate for equipment or a drive trace that is different than IM240. Most states with emissions test different than IM240 use software-based conversion factors to convert raw emissions results into EPA IM240 standard gpm. Using conversion factors to convert raw emissions readings to EPA IM240 standard grams per mile is similar to scoring standardized tests, like the MCAS and the SAT. In those tests, a raw score is obtained, the raw score is converted into a standardized score, and the standardized score is reported as the result of the test.

"EPA IM240 standard gpm" - Grams per mile of pollution emitted by a vehicle measured by IM240 equipment while running an IM240 drive trace. Almost all states with different equipment and drive traces scale (convert) the raw emissions to EPA standard IM240 gpm.

"Drive Trace" – The prescribed mix of acceleration, cruising, braking and speed that a vehicle is driven through when conducting an emissions test. Drive traces can be lower or higher speed, can be longer or shorter and can include variations in speed or constant speeds. Each different test has its own different drive trace.

"SIP" – The State Implementation Plan, or SIP, is a document a state submits to EPA which contains an inventory of the amount of pollution in a state, a description of the air pollution reductions needed to meet federal standards, and a list of the pollution reductions programs that a state will use to meet those federal standards. A SIP can describe a single program (e.g., I&M SIP) or describe a comprehensive plan to meet a federal standard (e.g., attainment SIP).

Test Reliability

"Errors of Commission" – Occurs when a vehicle failed a specific test, but should have passed its reference test. The reference test for MA31 is the IM240 equipment and drive trace. Therefore, Errors of Commission in Massachusetts refers to the failure of a vehicle using the MA31 equipment and drive trace but should have passed if tested on the IM240 equipment and drive trace.

"False failure" - Occurs when a vehicle failed a specific test but should not have. Every test has false failures. In Massachusetts, a "false failure" occurs when a vehicle fails MA31, but should have passed. Contributing factors can include vehicle variance, how the test is performed, or the equipment not working as designed.

"Vehicle variance" – The tendency of a motor vehicle to produce different amounts of pollution even when driven in exactly the same way. In I&M tests, vehicle variance is one of the reasons why the same vehicle can yield different test results from multiple tests.

Vehicle variance of clean cars tends not to affect test results, because the range of pollution emitted in typically well below the cutpoints. Vehicle variance of grossly polluting cars also tends not to affect test results, because the range of variability is typically well above the cutpoints.

However, vehicle variance from cars that pollute around the cutpoints tends to influence test results: they can pass one day and fail the next, even if two tests are performed in exactly the same way.

(NB: Vehicles that pollute around the cutpoints and the grossly polluting vehicles both have malfunctioning emissions control systems that cause them to pollute much more than they were designed to. Repairs to these vehicles are the way that I&M programs reduce pollution).

Test Effectiveness

"Excess emissions" - Pollution reductions that result from repairs of vehicles which fail an I&M test. Repairs made before the test are not counted as excess emissions, because the test did not identify them. As False Failures and Errors of Commission decrease, excess emissions also decrease, since 1) fewer vehicles are being failed and repaired; and 2) typically the vehicles identified in those categories show significant decreases in emissions after repairs are done. When analyzing test effectiveness – how well the equipment and drive trace perform in controlled circumstances – EPA assumes that vehicle pollution is reduced only to the cutpoints. When analyzing program effectiveness – how well the program works in the field – EPA analyzes data that show the actual reduction achieved from repairs.

"Program effectiveness" - Examines how well the equipment and trace perform as they are used in the field, and is usually analyzed as a part of a full program evaluation.

"Test effectiveness" – An EPA-required calculation expressed in percentages (e.g, 85%) that indicates how well a state's I&M test equipment and drive trace identify excess emissions from a vehicle fleet compared to EPA's "gold standard" test equipment and drive trace, the IM240. Analysis of test effectiveness is done under controlled conditions in order to isolate the performance of the equipment and the trace perform from the way they are used in an actual testing environment.

Descriptions of I&M Tests

Cutpoints – The pass/fail level for I&M tests. Vehicles which pollute above the cutpoints at the time they are tested fail. Vehicles which pollute below the cutpoints at the time they are tested pass. Different cutpoints are assigned for different pollutants.

IM240 test - EPA's "gold standard" equipment and drive trace for I&M programs. The "240" refers to the 240 second-by-second measurement of a vehicle emissions taken over the 239 second test. IM240 was derived from a small portion of the FTP (Federal Test Procedure), a two-day comprehensive test used by the federal government to certify emission levels from new vehicles.

IM240 start-up cutpoints – EPA-assigned pass/fail levels that are more lenient than final cutpoints. The cutpoints vary based on the pollutant being measured, the classification of the motor vehicle (e.g., passenger car vs. light duty truck) and amount of pollution the vehicle was allowed to pollute when it was new (the age of the vehicle). The cutpoints are expressed in EPA IM240 standard gpm.

IM 240 final cutpoints – EPA assigned pass/fail levels that are stricter than start-up cutpoints. States can choose whether to use the EPA start-up or the EPA final cutpoints. The cutpoints vary based on the pollutant being measured, the classification of the motor vehicle (e.g., passenger car vs. light duty truck) and amount of pollution the vehicle was allowed to pollute when it was new (the age of the vehicle). The cutpoints are expressed in EPA standard IM240 gpm.

MA31 test – The Massachusetts I&M test, consisting of the MASS99 equipment and the MA 31 drive trace. The MA31 has less expensive, less accurate equipment and a different and shorter drive trace. These factors cause the MA test to be less accurate than its reference test, IM240

MA31 start-up cutpoints - DEP assigned pass/fail levels that were used when the Enhanced Emissions and Safety Test began in October 1999. These lenient cutpoints varied based on the pollutant being measured, the classification of the motor vehicle (e.g., passenger car vs. light duty truck) and amount of pollution the vehicle was allowed to pollute when it was new (the age of the vehicle). The cutpoints were expressed in EPA IM240 standard gpm.

MA 31 final cutpoints – DEP assigned pass/fail levels in use since April 2001. DEP lowered the program's cutpoints from start-up to final phase in a series of steps between October 1999 and April 2001. **MA31** final cutpoints are expressed in EPA IM240 standard gpms and are contained in the I&M program regulations (310 CMR 60.02). MA 31 final cutpoints match EPA's start-up cutpoints.

COMMONWEALTH OF MASSACHUSETTS
EXECUTIVE OFFICE OF ENVIRONMENTAL AFFAIRS
DEPARTMENT OF ENVIRONMENTAL PROTECTION

MA31 Conversion Factor Analysis and Interim Test Effectiveness Evaluation

Massachusetts Enhanced Emissions and Safety Test

July 21, 2003

ABSTRACT:

The Massachusetts Department of Environmental Protection commissioned this study to: 1) determine revised factors to convert MA31 test results to equivalent IM240 test results, 2) evaluate MA31 test effectiveness compared to IM240 to determine whether EPA emission reduction commitments are being met, and 3) if necessary, evaluate selected program enhancements for improving test effectiveness. The study included emission testing 612 vehicles in Arizona using combinations of MA31 and IM240 drive traces and MASS99 and IM240 equipment, and evaluating a previously tested 2% sample of Arizona vehicles for determining MA31 test effectiveness. Results and recommendations are included. The study concludes that DEP should examine methods of increasing NOx effectiveness.

TABLE OF CONTENTS

APPENDICES

APPENDIX A: I&M Program Cutpoints

APPENDIX B: Monte Carlo Simulation Report

APPENDIX C: Monte Carlo Simulation for predicting MA31 scores from IM240 scores

APPENDIX D: Mileage Accumulation Rates from EPA MOBILE 6.2 Users Guide

APPENDIX E: MA147 to IM240 Regressions

APPENDIX F: Monte Carlo Simulation for predicting MA147 scores from IM240 scores

1.0 SUMMARY

Vehicles (i.e., cars, trucks, and buses) contribute nearly half of the emissions that cause ozone pollution (i.e., smog) in Massachusetts. The Enhanced Emissions and Safety Test (IM) program is a significant component of Massachusetts' State Implementation Plan (SIP) to control ozone pollution. The program reduces pollution from Massachusetts' vehicles by identifying those vehicles with seriously malfunctioning emissions controls and requiring that they be repaired. The program is designed to meet this objective at a reasonable cost and with a high degree of consumer convenience.

Modern IM programs test vehicles by using a dynamometer (i.e., a treadmill) to simulate actual driving conditions. An inspector drives the vehicle through a set protocol (or "drive trace") that includes a range of speeds and accelerations while an analyzer measures the amount of pollution the car produces per mile. The test identifies emissions of hydrocarbons (HC) and oxides of nitrogen (NOx), which combine to form ground-level ozone in the presence of sunlight, and of carbon monoxide (CO). The amount of each pollutant is then compared to a standard (called a "cutpoint") to determine whether the vehicle passes or fails.

The cutpoints are set so that the test fails enough of the dirtiest vehicles to reduce Massachusetts' ozone pollution that is attributable to motor vehicles to levels that fulfill the Commonwealth's commitments in its State Implementation Plan. The IM program, together with programs that reduce ozone precursors from other sources, has the goal of meeting federal clean air standards for ozone. Specific cutpoints have been set for each type of vehicle and model year, for each of the three pollutants of concern.

The U.S. Environmental Protection Agency (EPA) has developed standards for equipment, drive traces, and cutpoints in IM programs (referred to collectively as "IM240" after its 240-second drive trace). The IM240 laboratory-grade equipment is expensive and is therefore only cost effective in centralized IM programs with a few specialized inspection stations. EPA also allows states to establish decentralized programs, and to use different equipment and tests, as long as they can reliably identify most of the high-polluting vehicles that would be identified by IM240.

In the early 1990s Massachusetts made plans to implement EPA's model program: IM240 equipment in centralized test-only facilities. After similar programs were halted in other states, DEP and RMV held discussions with Massachusetts stakeholders (inspectors, repairers, motorists and officials from 11 state agencies), national I&M testing experts, and managers of I&M programs in other states. Based on these discussions, DEP and RMV began to develop a decentralized program.

After the Massachusetts Legislature enacted legislation in 1997 that authorized a decentralized program, DEP chose a shorter 31-second drive trace ("MA31") that reduces motorists' time waiting in line and limits the amount of noise in the shops (higher speed traces are much noisier). Massachusetts also chose less expensive test equipment

("MASS99") to minimize the cost of the test for motorists. This choice made program participation affordable for over 1500 inspection stations, thereby ensuring that Massachusetts motorists would have many convenient locations to choose from for their inspection. By comparison, Massachusetts would only have been able to afford to set up approximately 60 stations statewide using EPA's model program.

EPA approved the Massachusetts IM program in a Federal Register Notice published on November 16, 2000. Because Massachusetts does not use EPA's IM240 test, EPA's approval established specific interim "test effectiveness" targets for the Massachusetts test in its approval. These targets will be used with other information to calculate the level of pollution reduction for which the Massachusetts test can take credit. These test effectiveness targets were expressed as minimum percentages of excess emissions identified by the MA31 test as compared to the excess emissions identified by the IM240 test (excess emissions are the emissions reductions from repairs of vehicles that fail the test). EPA's test effectiveness targets for Massachusetts were established as: 85% for HC, 87% for CO, and 85% for NOx. EPA's approval acknowledges that the Massachusetts program is adequate to meet its goals for reducing pollution from the Commonwealth's vehicle fleet, although it will not identify all the excess emissions that would be identified by the IM240 test.

To determine whether the Massachusetts program is meeting these test effectiveness targets, EPA required DEP to perform a study evaluating the effectiveness of the Massachusetts IM test and equipment relative to EPA's "benchmark" IM240 test and equipment. In October 2000, DEP contracted with Sierra Research of Sacramento, California to design and perform the evaluation of Massachusetts' MA31 drive trace and MASS99 emissions test equipment. Sierra subcontracted with Gordon-Darby, Inc. to perform side-by-side MA31 and IM240 tests on a sample of vehicles using both the Massachusetts' MASS99 equipment and Gordon-Darby's IM240 test equipment. Gordon-Darby operates the Phoenix, Arizona program, which uses the IM 240 equipment and test, and is considered by EPA staff to be the best representation of EPA's "model" IM240 program.

Objectives. The study had three objectives:

1. Evaluate the conversion factors used by the Massachusetts program to express test results in terms that are comparable to IM240, and recommend improvements, if needed, so that Massachusetts test results more closely approximate results that would be expected from the IM240 test

2. Evaluate the effectiveness of the MA31 drive trace and equipment relative to the IM240 drive trace and equipment. "Test Effectiveness" measures how well a state's I&M test equipment and drive trace identify excess emissions from a vehicle fleet compared to EPA's benchmark equipment and drive trace, IM240.

3. Evaluate potential program changes that could increase the MA31 test effectiveness, if needed.

Conversion Factor Results. Sierra and Gordon-Darby tested 612 vehicles on both the MASS99 and IM240 test systems. The test data were first used to develop new conversion factors that express Massachusetts test results in terms that more accurately approximate IM240 test results than the initial conversion factors used by the program when it began in 1999. The program's initial conversion factors were based on studies from other states, which were the best data available at the time. In April 2001, Sierra and Gordon-Darby provided DEP with a preliminary analysis of 341 tests, which indicated that the initial conversion factors had been overestimating vehicle emissions compared to IM240, and were therefore failing more vehicles than was necessary to meet the Massachusetts goal for ozone reduction.

Based on this preliminary analysis, DEP determined that the conversion factors needed to be revised, and implemented revised conversion factors on July 11, 2001. The results of this change were: (1) the MA31 test more accurately simulated the IM240 test it was designed to mimic, and (2) emission levels and failure rates from the MA31 test decreased from their previous levels. The contractors continued to collect data through August 2001. The analysis of the full data set resulted in recommendations for additional modifications of the conversion factors.

Test Effectiveness Evaluation Results: The evaluation of the effectiveness of the MA31 test showed that, after the correction of the conversion factors in July 2001, it was exceeding target levels for HC and CO, and was not effective enough for NOx:

- HC: 87% compared to 85% target
- CO: 90% compared to 87% target
- NOx: 69% compared to 85% target.

Subsequent analyses investigated changes that could be made to the MA31 test to make it more effective. One change reduces the number of chances to pass the test from six drive traces in the initial program design to two (initially, if a vehicle passed the test in any one of the six, then it passed; this opportunity was known as "Fast Pass"). By allowing vehicles only two opportunities to pass, DEP's contractor showed that the test's effectiveness would be increased as follows:

- HC: 91% compared to 85% target
- CO: 93% compared to 87% target
- NOx: 75% compared to 85% target

DEP eliminated "Fast Pass" from the Massachusetts IM program on March 14, 2003.This change has allowed the Massachusetts test to exceed its targets for HC and CO, and to obtain 90% of the target for NOx.

Next Steps:

Improving Test Effectiveness: There are a number of ways in which the MA 31 test can be changed to more effectively identify excess NOx emissions. These include implementing a new drive trace for all vehicles to be tested on the dynamometer that is more similar to the IM240 ("MA 147"); selectively implementing MA 147 solely for vehicles that MA 31 does not identify as clearly clean; or changing the pass/fail cutpoints to make MA 31 a stricter test.

Program Evaluation: This report describes a study that looked at the performance of the MA31 drive trace and equipment under controlled conditions, and at the factors used to convert emissions results into terms that are comparable to the results of EPA's benchmark IM240 test. DEP will use the data produced by this study as well as information about how the Massachusetts program has worked in the field, to perform a broader evaluation of the program. This "phase 2" evaluation is also expected to result in recommendations for ways in which the program can be improved.

DEP will be discussing options for program modifications with EPA and other stakeholders. DEP is reviewing the potential effects of improvements in the MA 31 test design, including the expected result of implementing On Board Diagnostics (OBD) testing for virtually all 1996 and newer vehicles, to determine whether these program changes will result in Massachusetts meeting its target test effectiveness for NOx. In considering any program changes, DEP will seek to obtain the needed pollution reductions in a way that continues to balance the program's three goals: ensure convenience for motorists in terms of location and price; fits well with the private businesses that provide inspection and repair services; and achieves Massachusetts pollution reduction goals.

2.0 INTRODUCTION

2.1 Background

Motor vehicles are a predominant source of air pollution in Massachusetts. Based on an inventory of emissions for 1999, Massachusetts' Department of Environmental Protection (DEP) determined that on-road vehicles (i.e. passenger cars, trucks, and buses) account for nearly 50% of the ozone-creating pollutants hydrocarbon (HC) and oxides of nitrogen (NOx), and 60% of carbon monoxide (CO) emissions in the Commonwealth. According to EPA, one of the most effective ways to reduce motor vehicle emissions is to implement a vehicle inspection and maintenance (IM) program. An I&M program requires vehicles to receive periodic emissions tests and that vehicles which fail the test be repaired to reduce their emissions.

All new vehicle models sold in the United States must pass an EPA-approved emissions test called the Federal Test Procedure (FTP). Because the FTP drive cycle is lengthy (over 30 minutes) and the equipment is prohibitively expensive, it is not practical to use in an I&M program. EPA recognized this and recommended that I&M programs use a shorter 240-second test called the IM240 that is a subset of the FTP drive cycle. To allow for normal in-use degradation of the vehicle's emission control system, EPA established pass/fail cutpoints for the IM240 test that are at least two to three times higher (less stringent) than the FTP standards for new vehicles.

To assist states with their I&M program designs, EPA defined a "model" enhanced I&M program which includes the following elements:

- centralized network of test-only facilities,
- laboratory-grade IM240 emissions test equipment,
- IM240 drive trace, and
- EPA IM240 "final" or "start-up" cutpoints for determining passing and failing vehicles.

States that implement the IM240 test as designed in EPA's model IM240 program are given 100% credit for their I&M program's effectiveness at identifying excess HC, CO, and NOx emissions from vehicles.

Realizing that the "model" IM240 program is not practical or even necessary for all states to implement, EPA allows states flexibility in choosing alternative program designs. If states choose an alternative I&M program, however, they must define the relative effectiveness of the program in terms of percentage of excess emissions identified when compared to the model IM240 program. For Massachusetts, the MA31 test effectiveness is measured relative to the model IM240 program using EPA start-up cutpoints because that is what the MA31 test was designed to simulate.

The best way to determine the effectiveness of an alternative I&M program is to perform a study that directly compares the test traces and equipment from the two programs with side-by-side tests. In many instances, states need to implement the alternative I&M program before this study can be performed. In these cases, EPA has granted approval of the state's I&M program with an agreed upon level of program effectiveness and the contingency that a study would be performed to determine the *actual* test effectiveness.

2.2 The Massachusetts Enhanced Emissions and Safety Test

The Massachusetts Enhanced Emissions and Safety Test (I&M program) began in October 1999. This program includes the following major elements:

- a decentralized network of independent inspection stations,
- repair-shop grade emissions test equipment,
- a 31 second drive trace called the "MA31", and
- EPA's IM240 "start-up" cutpoints for determining passing and failing vehicles.

Keating Technologies (now Agbar Technologies) is the network contractor Massachusetts hired to operate the I&M program. The decentralized program network consists of approximately 1,500 stations spread throughout the Commonwealth. Agbar subcontracted with Environmental Systems Products, Inc (ESP) and the SPX Corporation (SPX) to provide the emissions test equipment (referred to as MASS99) and equipment repair services to the stations. The MA31 drive trace is the same as the "BAR31" drive trace that was developed by California's Bureau of Automotive Repair (BAR) and is currently used in the Oregon and Rhode Island I&M programs. Finally, DEP chose to use EPA's IM240 "start-up" cutpoints (instead of the more stringent "final" cutpoints) because these were sufficient to meet the Commonwealth's emission reduction goals.

In the Massachusetts I&M program, the MASS99 equipment calculates emissions from the MA31 trace and converts them to equivalent IM240 scores using a separate conversion factor for each of the three pollutants, HC, CO, and NOx. These conversion factors are designed to account for the differences between the MA31 and IM240 trace *and* the differences between the MASS99 equipment and the IM240 equipment.

At the beginning of Massachusetts' I&M program, conversion factors of 1.5, 0.86, and 0.86 were used for HC, CO, and, NOx, respectively. These conversion factors were derived from a study of New York's NYTEST I&M program, which uses the IM240 drive trace and the same equipment as the Massachusetts I&M program. These conversion factors account for the differences between the MASS99 and IM240 equipment, but not the differences between the MA31 and IM240 drive traces. The differences between the MA31 and IM240 traces were accounted for by using BAR31 to IM240 drive trace conversion factors developed by Oregon's I&M program. However, instead of applying these conversion factors to the raw MA31 test scores (like the NYTEST conversion factors), these conversion factors were applied to the Massachusetts cutpoints during program start-up. As the cutpoints were ratcheted down to their final values (EPA IM240 start-up cutpoints) in April 2001, it became necessary to establish new MA31 to IM240 conversion factors that accounted for both the difference in equipment and drive trace.

Due to the differences between Massachusetts' MA31 test and MASS99 equipment and EPA's "model" IM240 program, EPA set the MA31 test effectiveness at 85%, 87%, and 85% for HC, CO, and NOx, respectively, when compared to the IM240 test with "start-up" cutpoints. These test effectiveness values were derived in part from New York's study of their NYTEST program, which uses the same test equipment as Massachusetts. DEP was required to use these values in EPA's MOBILE6 emission factor model to determine on-road vehicle emissions for Massachusetts' State Implementation Plan (SIP). As part of the SIP submittal to EPA, Massachusetts committed to performing an evaluation of its test trace and equipment to determine the actual MA31 test effectiveness compared to EPA's "model" IM240 program.

3.0 STUDY DESIGN

DEP contracted with Sierra Research of Sacramento, California (Sierra) to perform the study. Sierra subcontracted with Gordon-Darby, Inc of Louisville, KY (Gordon-Darby) to collect IM240 and MA31 test data on up to 1,000 vehicles over a range of vehicle types, model years, and emission rates. Gordon-Darby operates the emissions testing program for the greater Phoenix, Arizona area. That Arizona I&M program is widely considered to best represent EPA's "model" I&M program consisting of centralized testing, IM240 drive trace, and IM240 laboratory grade equipment.

As shown in Table 1, the study required emissions data from both the IM240 and MA31 drive traces to be collected on both Gordon-Darby's IM240 test equipment and the MASS99 test system. Each drive trace was run on each type of equipment with back-to-back tests for each of the test vehicles. Each test combination has been assigned a name (e.g. "MA240" is the IM240 drive trace driven on MASS99 test equipment) that is referenced throughout this report.

Table 1
Study Design with Designated Test Names

	Drive Trace	
Equipment	IM240	MA31
Gordon-Darby IM240	IM240 test	IM31 test
MASS99	MA240 test	MA31 test

Agbar Technologies (Agbar), Massachusetts' contractor for operating the program in the state, was responsible for providing the MASS99 test system and maintenance during the study. Agbar provided a MASS99 system manufactured by SPX Corporation, one of the two equipment providers in the Massachusetts program, for the study.

3.1 Vehicle Selection

One of the primary goals of the study was to develop conversion factors between emissions rates from the MA31 drive cycle using MASS99 test equipment and the IM240 test using IM240 test equipment. To develop these conversion factors, it was necessary to collect test data over the *entire* range of vehicle types, vehicle model years, and vehicle emission rates that are observed in the Massachusetts program, including older dirtier vehicles with high emissions. For this reason, it should be noted that the distribution of vehicles for the study is likely not representative of the vehicle fleet in the Massachusetts I&M program, which is weighted more heavily toward newer, lower-emitting vehicles.

Test vehicles for the study were recruited from vehicles arriving at a Gordon-Darby's test center for their normal Arizona emissions compliance inspection, the IM147 test. The IM147 test is a shortened version of the IM240 test, consisting of the second phase (high speed phase) of the IM240 test. Vehicles were chosen for inclusion in the study after they had completed their compliance inspection. Results from the IM147 test were used to sort and select vehicles for the study. Vehicle stratification for the study was established by vehicle type, model year, and emissions rates for HC, CO, and NOx based on the IM147 compliance test. This gave a distributed sample of vehicles with a total of 81 categories of vehicles, or "bins," for the data to fall into as shown in Table 2.

Table 2
Bins for Vehicle Selection

Vehicle Type	Model Years	HC (g/mi)			CO (g/mi)			NOx (g/mi)		
		Low	Mid	High	Low	Mid	High	Low	Mid	High
LDGV (17 per model year bin, 450 max)	81-84 85-89 90+	< 1.28	≥ 1.28 and < 2.55	≥ 2.55	< 30.4	≥ 30.4 and < 60.8	≥ 60.8	< 2.55	≥ 2.55 and < 5.09	≥ 5.09
LDGT1 (13 per model year bin, 350 max)	81-84 85-89 90+	< 2.08	≥ 2.08 and < 4.17	≥ 4.17	< 50.4	≥ 50.4 and <100.7	≥100.7	< 3.56	≥ 3.56 and < 7.13	≥ 7.13
LDGT2 (8 per model year bin, 200 max)	81-84 85-89 90+	< 2.76	≥ 2.76 and < 5.53	≥ 5.53	< 53.6	≥ 53.6 and <107.3	≥107.3	< 4.00	≥ 4.00 and < 7.99	≥ 7.99

LDGV = Light Duty Gasoline Vehicles 0 to 6,000 lbs. Gross Vehicle Weight Rating (GVWR)
LDGT1 = Light Duty Gasoline Trucks 0 to 6,000 lbs. GVWR
LDGT2 = Light Duty Gasoline Trucks 6,001 to 8,500 lbs. GVWR

The model year bins were chosen to generally group vehicles by fueling technology type (carbureted or fuel injected), based on data from EPA's MOBILE6 model. The three emissions rate categories were established and used to recruit low, medium, and high emitters for the study based on their IM147 compliance test results. The initial thresholds for low, medium and high emitters were developed based on an analysis of IM147 test data provided by Gordon-Darby that was collected in October 2000 at the same inspection station used in the study.

3.2 Test Equipment

A fully compliant IM240 test system at the Gordon-Darby test facility was used to perform the IM240 testing and was specially programmed to run the MA31 test cycles. The IM240 test system is comprised of an AC electric, full-inertia simulation, dual 8.65" roll chassis dynamometer; a constant volume sampler system; and analyzers compliant with the EPA IM240 requirements.

The Commonwealth contracted with Agbar to place one production MASS99 workstation at Gordon-Darby's facility. The MASS99 workstation was programmed to run both the MA31 and MA240 tests. Agbar chose to use a MASS99 workstation provided by SPX, one of Agbar's two equipment suppliers for the Massachusetts program. Agbar was responsible for getting the analyzer installed and operational, working with Gordon-Darby to train their inspectors on the operation of the test system (including methods for removing the test data from the system), and maintenance of the test system during the study.

All equipment during the study was operated by Gordon-Darby staff in a special test lane that was dedicated to the study. The test equipment was set up so that vehicles were first tested on the IM240 equipment, followed by the MASS99 system.

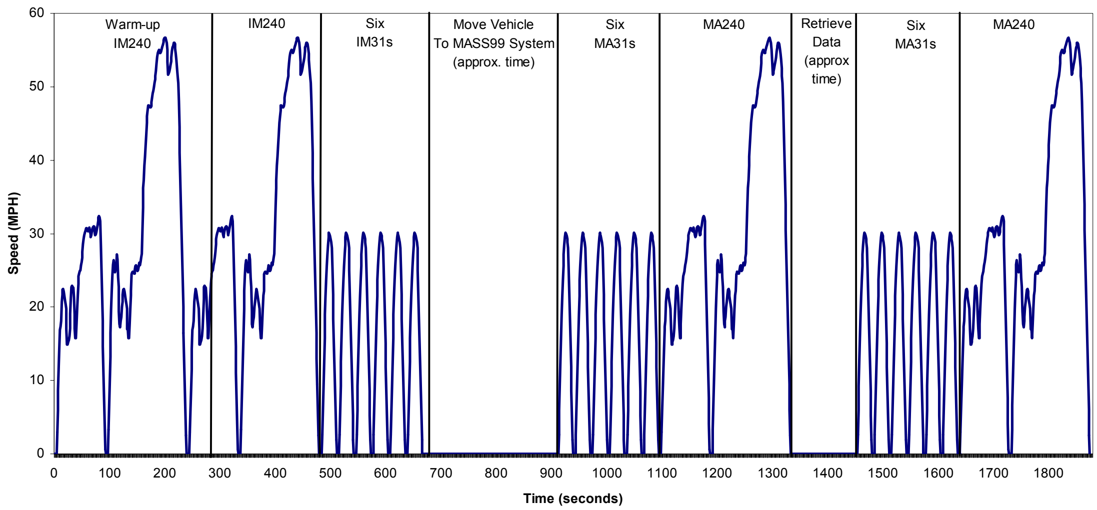
3.3 Test Protocol

Each test vehicle received its regular compliance test (the IM147) as required in the Arizona I&M program. The compliance test has built-in retest algorithms to ensure the vehicle is properly preconditioned. If the vehicle fell into one of the testing bins described previously, the motorist was offered an incentive to participate in the study. If the motorist agreed, the vehicle was directed into the special study test lane and placed on the IM240 dynamometer and connected to the IM240 analytical system.

Figure 1 presents the original drive cycle protocol for the study. In the study test lane, the testing began with an IM240 test cycle to ensure the vehicle was warmed up. Although the vehicle is warmed up at the end of the compliance test, there was a chance the vehicle could cool slightly between the end of the compliance test and the start of the drive cycle for the study. Next, another full-duration IM240 test cycle was conducted, followed by six MA31 drive cycles (IM31). The vehicle was then moved to the MASS99 test system. The time between completing the testing on the IM240 test system and beginning testing on the MASS99 system was kept as short as possible to ensure that the vehicle remained at normal operating temperature. On the MASS99 test system, six MA31 drive cycles were conducted followed by one "MA240" (IM240 drive cycle driven on MASS99 test equipment). The sequence on the MASS99 equipment was then repeated, six MA31 drive cycles followed by one MA240.

Figure 1

Original Drive Cycle Protocol



3.4 Data Collection

For each test vehicle, specific vehicle identification information was collected as well as summary emission data for each test cycle and second-by-second data during the test. Table 3 shows the data that were collected for each test vehicle on the IM240 and MASS99 test systems.

Table 3
Test Data Collected

Vehicle Data	Emissions Test Data	
(collected for both test systems)	**IM240 Equipment**	**MASS99 Equipment**
VIN	Drive cycle number	Drive cycle number
License plate	HC (grams per mile)	HC (grams per mile)
Make	CO (grams per mile)	CO (grams per mile)
Model	CO2 (grams per mile)	CO2 (grams per mile)
Model year	NOx (grams per mile)	NOx (grams per mile)
Body style	Speed	Speed
Body Type	Dilution ratio	Dilution ratio
Transmission type	CVS flow rate	Seconds in test
Number of cylinders	Seconds in test	Dilute flow rate
GVWR (for trucks)		Dilute temperature
Odometer		Dilute pressure
EPA Sierra Lookup Table (ESLT) ID number		Dilute O2 concentration
Dynamometer power setting		Raw HC concentration
Test date		Raw CO concentration
Test time		Raw CO2 concentration
Inspector ID		Raw NO concentration
Ambient temperature		Raw O2 concentration
Relative humidity		Exhaust volume
Barometric pressure		

3.5 Quality Assurance

Gordon-Darby was responsible for the quality assurance of their test system and the data it collected. They were also responsible for collecting and archiving data generated by the MASS99 system.

Agbar (and its supplier, SPX) was responsible for the quality assurance of the MASS99 unit when it was installed in the test lane. Agbar/SPX trained Gordon-Darby staff on the necessary quality assurance procedures for the equipment (how to perform calibrations, etc.), which Gordon-Darby was jointly responsible for during the study. Agbar and SPX were responsible for maintaining the MASS99 unit and providing technical support for operational issues.

When the MASS99 unit was initially installed, Sierra was on-site for one week to perform an audit of the MASS99 system (dynamometer and gas analyzer) to ensure it was functioning properly. This was similar to Massachusetts' overt audit procedure for MASS99 analyzers, but only included those items used in the correlation study (e.g., the gas cap tester was not audited). Calibration information for the Gordon-Darby IM240 test system was reviewed by Sierra to ensure that it was in good operating condition and had been properly calibrated. All applicable drive cycles were tested to ensure they were programmed properly, and the data from both the Gordon-Darby system and the MASS99 system were checked to ensure all of the required information was being properly collected. Sierra assisted in resolving any problems involved in getting the test systems operational, and observed some of the initial vehicle selection, recruitment, and testing to make sure the test protocols were being properly followed.

While the testing progressed, Sierra received test data from the program, performed ongoing data quality control checks, and set up the tools necessary to analyze the test data. The actual distribution of test vehicles was evaluated by Sierra against the target distribution to determine if the established sample bins were being properly filled. Sierra also visited the Gordon-Darby test lane several times during the data collection phase to manually collect data, recheck the performance of the test protocol, and coordinate changes to the vehicle selection criteria as needed.

4.0 VEHICLE TESTING AND DATA COLLECTION

The following sections describe vehicle testing, data collection, and changes to the study design that occurred during testing.

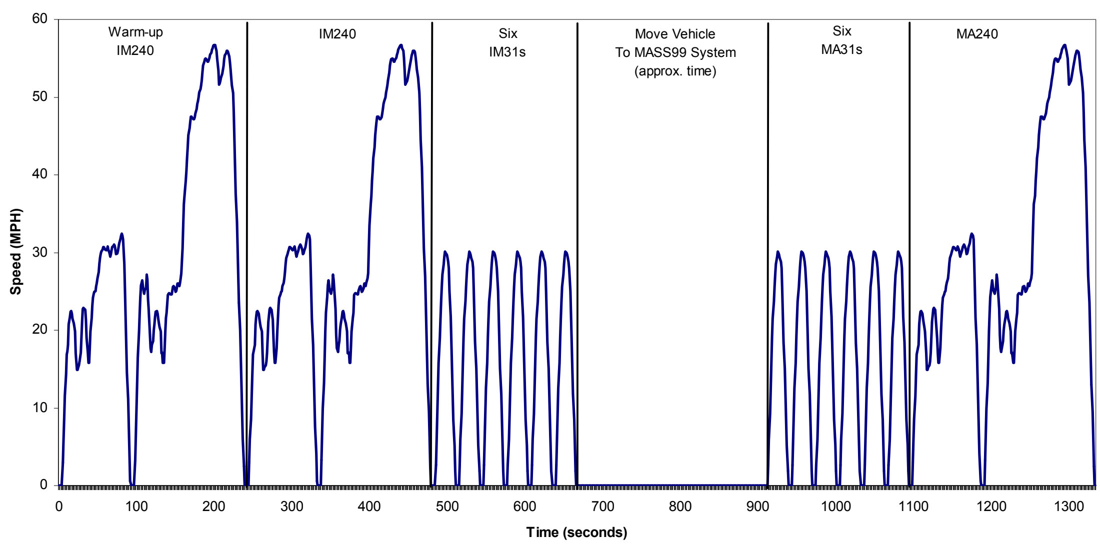
4.1 Vehicle Testing

Testing was performed from December 12, 2000 to August 18, 2001 in the Gordon-Darby test lane in Arizona. Roughly 850 vehicles were tested in the study.

In February 2001, the drive cycle for the study was modified to eliminate repeating the second test sequence on the MASS99 equipment (see Figure 2). These additional test cycles were eliminated early in the data collection phase because (1) sufficient data had been collected to allow evaluation of test-to-test variability, (2) the extended test protocol was causing some vehicle engine overheating problems, and (3) the need to collect the data off the SPX system between test modes was extending total test time beyond that acceptable to most motorists. Subsequent analysis of the full dataset utilized the first six MA31 cycles and the first MA240 cycle, since these data were collected from all test vehicles in the study.

Figure 2

Modified Drive Cycle Protocol



During the testing period, Sierra adjusted limits for its low, medium, and high emissions bins due to difficulty in locating and recruiting high emitting vehicles for the study. Table 4 shows the emissions limits for the bins at the end of the study.

Table 4
Final Vehicle Emissions Bins

Vehicle Type	Model Years	HC (g/mi)			CO (g/mi)			NOx (g/mi)		
		Low	Mid	High	Low	Mid	High	Low	Mid	High
LDGV	81-84 85-89 90+	< 1.0	≥ 1.0 and < 2.0	≥ 2.0	< 15	≥ 15 and < 30	≥ 30	< 1.5	≥ 1.5 and < 3.0	≥ 3.0
LDGT1	81-84 85-89 90+	< 1.5	≥ 1.5 and < 3.0	≥ 3.0	< 30	≥ 30 and < 60	≥ 60	< 2.5	≥ 2.5 and < 5.0	≥ 5.0
LDGT2	81-84 85-89 90+	< 2.0	≥ 2.0 and < 4.0	≥ 4.0	< 30	≥ 30 and < 60	≥ 60	< 2.5	≥ 2.5 and < 5.0	≥ 5.0

4.2 Data Collection

Data from the study were collected from two different sets of equipment, the MASS99 system manufactured by SPX and the Gordon Darby IM240 system. Data from the MASS99 system were recorded on three separate files. One file contained summary data for the inspection such as the 17 digit Vehicle Identification Number (VIN) vehicle information, and the overall test results. Another file contained summary results from the different cycles of the inspection. Finally, the third file contained second-by-second results from all of the cycles. Data from the IM240 system were recorded in a single file.

Due to periodic problems with data file corruption with the MASS99 system and errors entering the 17 digit VIN into the two systems, the number of valid test records was reduced to 612. Tables 5, 6, and 7 below show the breakdown of the 612 tests by vehicle type, model year, pollutant, and emission rates, based on emissions from the AZ IM147 compliance test used to recruit the vehicles.

Table 5
Number of LDGVs Tested by Model Year, and Emissions Rates

Model Year	HC (g/mi)			CO (g/mi)			NOx (g/mi)		
	Low (< 1.0)	Mid	High (≥ 2.0)	Low (< 15)	Mid	High (≥ 30)	Low (< 1.5)	Mid	High (≥ 3.0)
81-84	33	5	7	31	11	3	20	17	8
85-89	73	18	14	77	15	13	54	35	16
90+	156	16	4	158	12	6	124	38	14
All	262	39	25	266	38	22	198	90	38

Table 6
Number of LDGT1s Tested by Model Year, and Emissions Rates

Model Year	HC (g/mi)			CO (g/mi)			NOx (g/mi)		
	Low (< 1.5)	Mid	High (≥ 3.0)	Low (< 30)	Mid	High (≥ 60)	Low (< 2.5)	Mid	High (≥ 5.0)
81-84	12	6	7	14	6	5	15	8	2
85-89	63	23	12	81	9	8	52	36	10
90+	72	9	4	77	0	8	73	10	2
All	147	38	23	172	15	21	140	54	14

Table 7
Number of LDGT2s Tested by Model Year, and Emissions Rates

Model Year	HC (g/mi)			CO (g/mi)			NOx (g/mi)		
	Low (< 2.0)	Mid	High (≥ 4.0)	Low (< 30)	Mid	High (≥ 60)	Low (< 2.5)	Mid	High (≥ 5.0)
81-84	4	0	0	3	1	0	2	2	0
85-89	24	2	2	26	0	2	15	11	2
90+	38	6	2	44	2	0	27	14	5
All	66	8	4	73	3	2	44	27	7

5.0 CONVERSION FACTOR ANALYSES

Conversion factors are used by the MASS99 system to automatically convert raw MA31 scores to equivalent IM240 scores. This is necessary because the Massachusetts program uses pass/fail cutpoints established by EPA based on the IM240 test. In this study, conversion factors were developed using a linear regression with the raw MA31 score as the independent variable while the IM240 score was the dependent variable.

The format of the linear equation used in the MASS99 system is:

IM240 Equivalent Score = Raw MA31 Score x Conversion Factor

The MASS99 system compares the IM240 equivalent scores to the Massachusetts program cutpoints to determine if the vehicle passes or fails the MA31 test. The IM240 equivalent scores are the pollutant readings that appear on the vehicle inspection report given to the motorist.

5.1 Analysis of Partial AZ study Dataset

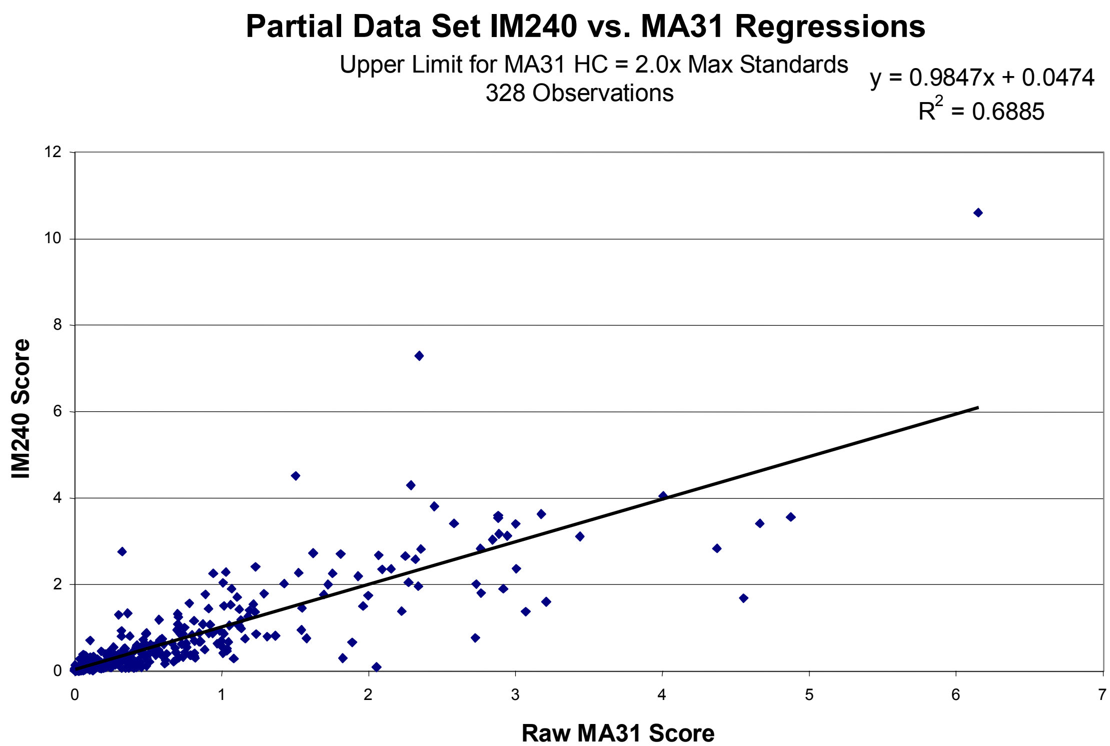
Prior to completing the testing in Arizona, DEP and Sierra used a partial dataset of 341 tests to calculate interim MA31 to IM240 conversion factors. DEP wanted to perform this interim analysis to determine if the initial conversion factors used in the Massachusetts program needed to be changed.

Since the emissions reported by the MASS99 system during the study were adjusted to IM240 equivalent scores using the initial conversion factors, the first step of the analysis was to convert back to raw MA31 scores by dividing by the initial conversion factors (see above equation). The initial conversion factors were 1.5, 0.86, and 0.86 for HC, CO, and NOx, respectively.

Since the main purpose of the conversion factors is to facilitate accurate pass/fail decisions relative to the IM240 cutpoints, it is essential that the regressions generating the conversion factors are accurate for vehicles performing near the cutpoints. As a result, some vehicles with higher emissions were removed from the sample to prevent them from disproportionately affecting the regression. To determine a reasonable threshold for removing high emission vehicles, a number of scatter plots were created. These plots ranged from including the entire dataset to a limited subset depending on emission cutpoints. After reviewing the plots, raw MA31 scores greater than 2.0 times the highest applicable cutpoint for all of the vehicles were removed from the interim regression analysis. A linear regression was chosen for the regression analysis because the MASS99 software is designed for a linear conversion of emission scores.

Figures 3 through 5 show regressions for vehicles having MA31 emissions less than or equal to 2.0 times the maximum MA31 cutpoint for each pollutant. The maximum MA31 cutpoints are 3.2, 80, and 7.0 grams per mile (g/mi) for HC, CO, and NOx, respectively. Each pollutant was considered independently for this analysis, which is why there are a different number of observations for each regression. For example, vehicles having HC emissions greater than 2.0 times the maximum HC cutpoint could appear on the CO figure, as long as its CO emissions were less than or equal to 2.0 times the maximum CO cutpoint.

Figure 3



For each pollutant, the regression analysis yielded a linear equation for converting raw MA31 scores to equivalent IM240 scores in the form of y = mx + b, where y is the IM240 score, x is the raw MA31 score, m represents the slope or conversion factor, and b is the y-intercept. The correlation coefficient (R^2) was also calculated, which expresses the relative strength of the association between x and y. An R^2 value of 1.0 would mean the two samples are perfectly correlated; i.e., all data points would lie exactly upon the regression line. Both the equation and the calculated correlation coefficients are shown on the charts.

Figure 4

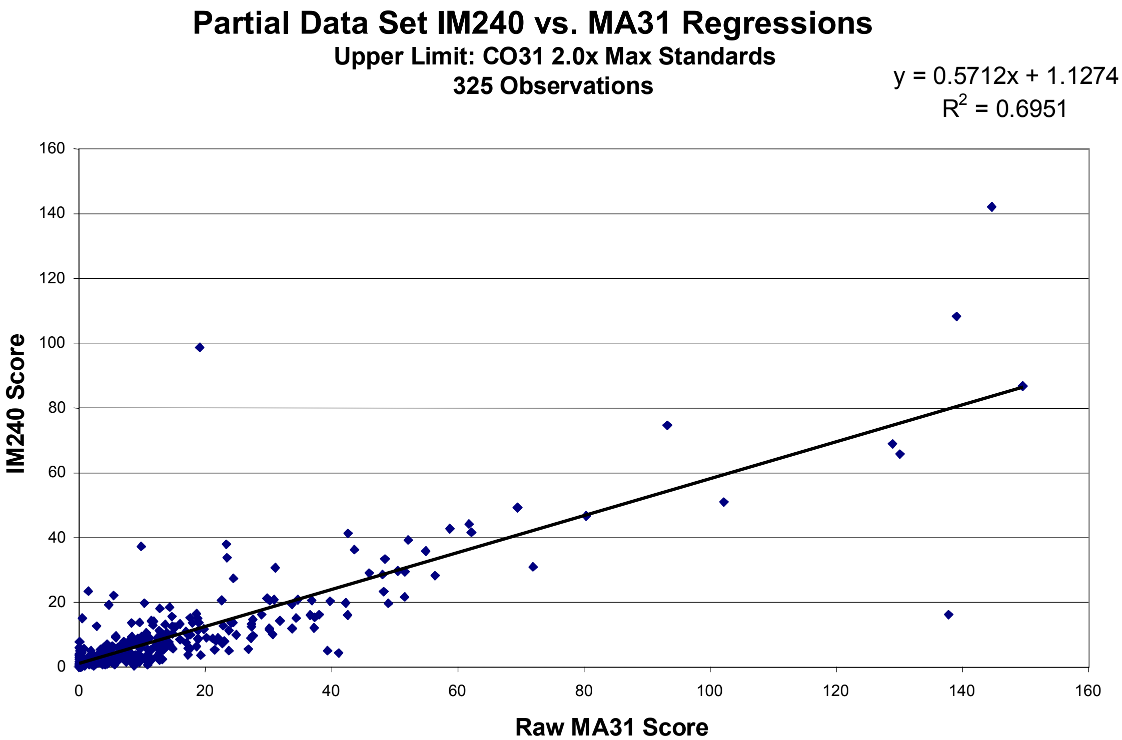


Figure 5

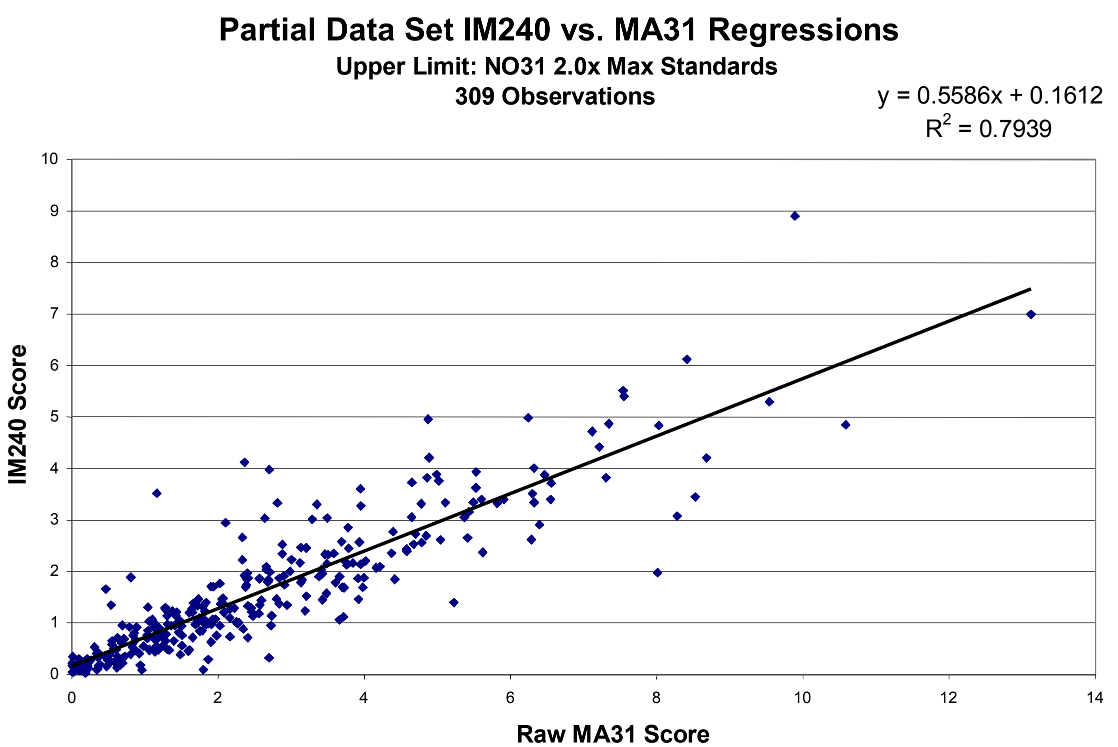


Table 8 compares these interim conversion factors to those initially used by the program.

Table 8

Comparison of Initial and Interim Conversion Factors

	HC	CO	NOx
Massachusetts I&M Program Initial	1.50	0.86	0.86
Interim: based on AZ 341 Sample Dataset	0.98	0.57	0.56

Analysis of these data show that the initial conversion factors were significantly higher than those calculated from the AZ study partial dataset. This means that, based on the AZ data, the conversion factors were causing the MASS99 equipment to overestimate IM240 emissions. At the beginning of Massachusetts' I&M program, cutpoints were set high (less stringent). However, by the time the final cutpoints (EPA IM240 start up cutpoints) were implemented in April 2001, the initial conversion factors were causing the transient test to be more stringent than was intended. Following this analysis, on July 11, 2001 the conversion factors were changed to the interim values calculated from the AZ 341 sample dataset. There were two outcomes from this change: 1) the MASS99 equipment was no longer overestimating IM240 emissions from the MA31 test, and 2) the reported emission scores and the number of emissions failures were reduced accordingly.

5.2 Analysis of Full AZ Study Dataset

Upon completion of the testing in August 2001, MA31 to IM240 conversion factors were calculated for the full 612 sample dataset. This regression analysis is the same as described above with the following exceptions: (1) only MA31 results less than or equal to 1.5 times the maximum MA31 cutpoint were used in the regressions so as to exclude more datapoints well above the cutpoint and (2) the regression equations were "forced" through zero (i.e. $y = mx$), eliminating the y-intercept to match the form of the equation used by the MASS99 software.

Figures 6 through 8 show regressions for the three pollutants based on the full 612-sample dataset.

Figure 6

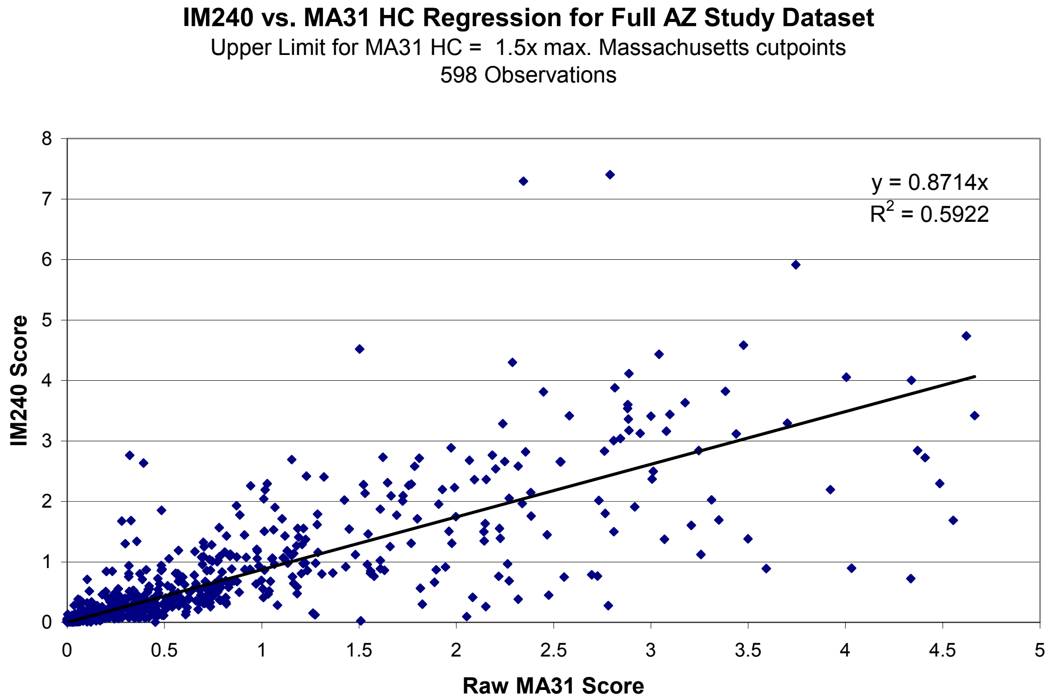


Figure 7

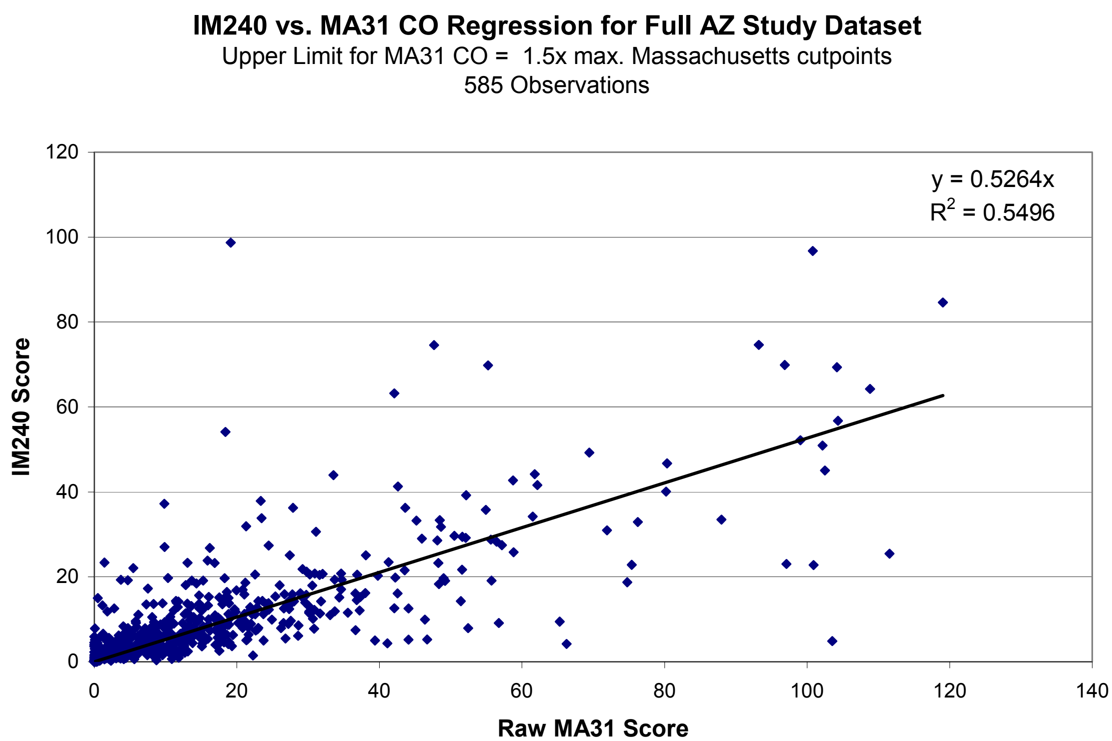


Figure 8

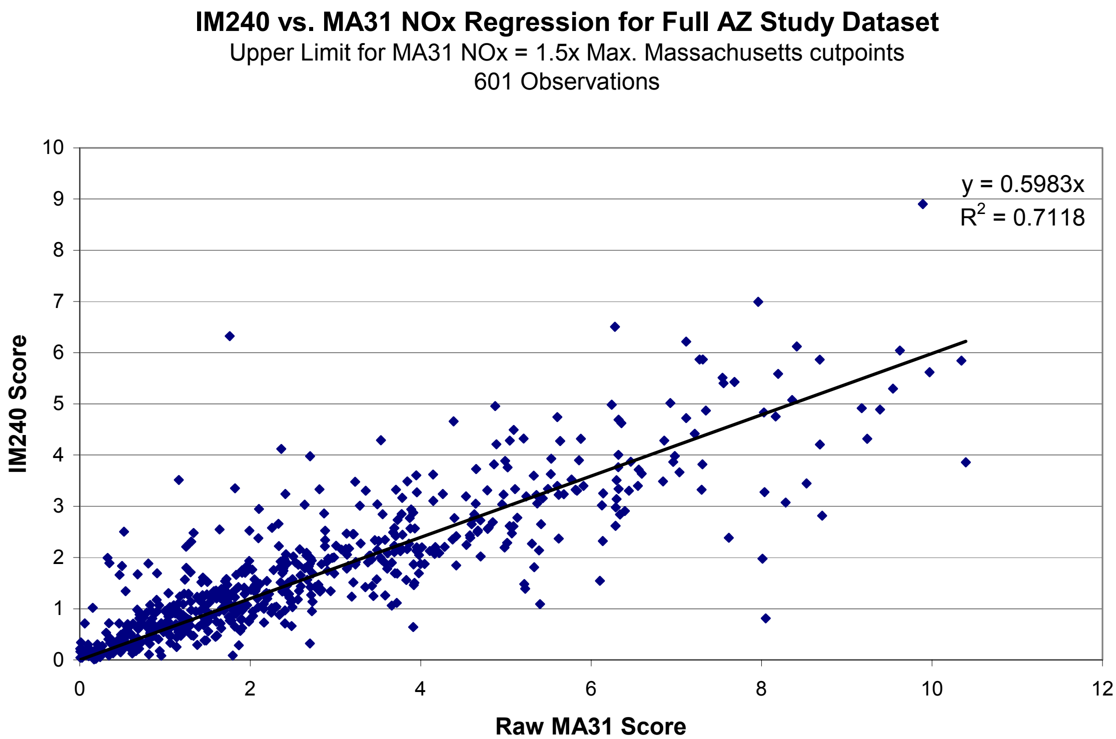


Table 9 compares conversion factors developed from both the full and partial datasets to the initial factors used in the program.

Table 9
Comparison of All Conversion Factors

	HC	CO	NOx
Massachusetts I&M Program Initial	1.50	0.86	0.86
Interim, implemented July 2001 (based on AZ 341 Sample Dataset)	0.98	0.57	0.56
Final (based on AZ 612 Sample Dataset)	0.87	0.53	0.60

As the data indicate, the conversion factors developed from the 612 sample AZ study dataset were fairly close to the interim conversion factors implemented in July 2001, with HC and CO slightly lower and NOx slightly higher. Because these results were similar, DEP did not want to implement the final conversion factors until the program effectiveness analyses were completed and checked.

6.0 METHODOLOGY FOR EVALUATING TEST EFFECTIVENESS

For these and subsequent analyses in this report, test effectiveness was evaluated by Sierra in terms of failure rates, errors of commission (EOC), and excess emissions identified by the MA31 transient test. These three parameters are commonly used to assess the relative efficiency of alternative I&M tests. These analyses were first performed on the full 612-sample dataset from the AZ study in Section 6.1. However, because the AZ study dataset was designed to be biased toward higher emitting vehicles (to benefit the conversion factor analysis), it was necessary to perform further evaluation (Section 6.2) to make these results relevant to the Massachusetts vehicle fleet. To do this, a statistical method of predicting realistic MA31 scores from IM240 scores was developed using data from the AZ study and a "Monte Carlo" simulation as presented in Section 6.2.1. This statistical method was then used in Section 6.2.2 to determine the effectiveness of the Massachusetts MA31 test using a dataset of IM240 tests that better represented the characteristics of the Massachusetts vehicle fleet.

The analyses in this section all assume the Massachusetts test sequence consists of six MA31 test traces with one chance to pass on the final trace. Also, the analyses uses the final MA31 to IM240 conversion factors calculated from the full AZ study dataset in the previous section.

6.1 Using the AZ Study Dataset

6.1.1 Failure Rates

Failure rates for the full 612-vehicle AZ study dataset were calculated using the final conversion factors determined in the previous section. To perform this analysis, IM240 equivalent scores (final MA31 scores) were first divided by the initial conversion factors used by the MASS99 equipment in the AZ correlation study. The resulting raw MA31 scores were then multiplied by the final conversion factors to obtain new IM240 equivalent scores from the MASS99 equipment. These scores were then compared to the final MA31 cutpoints implemented in the Massachusetts program to determine the failure rate. These MA31 cutpoints are the same as EPA's IM240 start up cutpoints. Appendix A presents the different cutpoints relevant to these and subsequent analyses.

The IM240 start up failure rate was calculated by comparing the IM240 scores generated by the Gordon-Darby IM240 equipment to EPA's IM240 start up cutpoints. The IM240 final failure rate was calculated by comparing the same IM240 scores from the study to EPA's IM240 final cutpoints. Table 10 presents the results from these failure rate analyses.

Table 10
MA31 Failure Rates
AZ Study Dataset – 612 Vehicles

Vehicle Type	Model Years	Vehicle Count	MA31 Failure Rates				IM240 Failure Rates	
			HC	CO	NOx	Overall	Start-Up Cutpoints	Final Cutpoints
LDGV	81 - 84	45	6.7 %	4.4 %	15.6 %	22.2 %	28.9 %	62.2 %
	85 - 89	105	10.5 %	11.4 %	17.1 %	31.4 %	28.6 %	57.1 %
	90 +	176	9.1 %	6.3 %	10.8 %	21.6 %	15.3 %	29.6 %
	All	326	9.2 %	7.7 %	13.5 %	24.8 %	21.5 %	42.9 %
LDGT1	81 - 84	25	8.0 %	8.0 %	0.0 %	16.0 %	24.0 %	52.0 %
	85 - 89	98	8.2 %	3.1 %	8.2 %	18.4 %	21.4 %	51.0 %
	90 +	85	3.5 %	3.5 %	9.4 %	14.1 %	18.8 %	23.5 %
	All	208	6.3 %	3.8 %	7.7 %	16.3 %	20.7 %	39.9 %
LDGT2	81 - 84	4	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	85 - 89	28	7.1 %	3.6 %	7.1 %	14.3 %	14.3 %	32.1 %
	90 +	46	17.4 %	0.0 %	13.0 %	23.9 %	21.7 %	30.4 %
	All	78	12.8 %	1.3 %	10.3 %	19.2 %	18.0 %	29.5 %
Total	All	612	8.7 %	5.6 %	11.1%	**21.2 %**	**20.8 %**	**40.2 %**
Note: These failure rates are for the Arizona study dataset only, which was not intended to represent the distribution of vehicles in the Massachusetts fleet. Also, since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates.								

The data show that the overall MA31 failure rate is approximately the same as the IM240 startup failure rate and substantially lower than the IM240 final failure rate. Since the MA31 test is designed to mimic IM240 with start up cutpoints, the focus of the comparison should be on those two sets of results.

6.1.2 Errors of Commission

Errors of commission (EOC – also referred to as "false failures") occur when a vehicle that would have passed a reference test, in this case the IM240 test, fails an alternative test, in this instance, the MA31 test. Measurement of EOC is primarily a tool for estimating the relative precision and efficiency of I&M test options such as those considered in Section 9. EOC is evaluated to ensure that a particular test procedure does not result in an excessive number of vehicles being failed that would not fail a reference test (i.e., the IM240 in this case). Some options for increasing test effectiveness may also increase EOC, however others can increase test effectiveness while reducing or maintaining the level EOC. Measuring EOC helps indicate which changes are the most efficient changes.

EPA sets no standard for EOC rates because their primary concern is ensuring that a sufficient number of high emitters are identified. The majority of EOC occur with vehicles that are polluting in excess of their original certification standards and, when repaired, show reduced emissions that benefit air quality. Sierra considers an EOC rate of 5% of all tested vehicles relative to EPA IM240 final cutpoints to be a reasonable and acceptable EOC level when compared to other programs around the country.

Table 11 shows the MA31 EOC rates versus EPA IM240 with startup and final cutpoints by pollutant as well as the overall EOC rate for the 612-vehicle dataset. The EOC rates are calculated as a percentage of the total tests performed. Note that the overall EOC rate for each category may be less than the sum of the individual pollutant failure rates because a vehicle may fail more than one pollutant during the test. The overall EOC rate for this dataset relative to EPA final IM240 cutpoints is 2.6%, which is within the reasonable and acceptable range for a properly functioning I&M program.

Table 11
MA31 EOC Rates
AZ Study Dataset – 612 Vehicles

Vehicle Type	Model Year	Vehicle Count	EOC Rates vs. EPA IM240 Start-Up Cutpoints				EOC Rates vs. EPA IM240 Final Cutpoints			
			HC	CO	NOx	Overall	HC	CO	NOx	Overall
LDGV	81-84	45	0.0	0.0	8.9	8.9	0.0	0.0	0.0	0.0
	85-89	105	2.9	1.9	2.9	6.7	1.0	0.0	1.9	2.9
	90+	176	5.1	3.4	3.4	9.7	2.8	1.1	0.6	4.0
	Total	326	3.7	2.5	4.0	8.6	1.8	0.6	0.9	3.1
LDGT1	81-84	25	4.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0
	85-89	98	3.1	1.0	4.1	8.2	0.0	0.0	2.0	2.0
	90+	85	1.2	0.0	1.2	2.4	1.2	0.0	1.2	2.4
	Total	208	2.4	0.5	2.4	5.3	0.5	0.0	1.4	1.9
LDGT2	81-84	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	85-89	28	3.6	0.0	0.0	3.6	3.6	0.0	0.0	3.6
	90+	40	4.3	0.0	2.2	6.5	0.0	0.0	2.2	2.2
	Total	78	3.8	0.0	1.3	5.1	1.3	0.0	1.3	2.6
Grand Total		612	3.3	1.5	3.1	**7.0**	1.3	0.3	1.1	**2.6**

Note: These EOC rates are for the Arizona study dataset only, which was not intended to represent the distribution of vehicles in the Massachusetts fleet.

EOC occur in test effectiveness studies for 3 reasons:

1. vehicles vary in the emissions they produce – that is, vehicles do not emit at a constant rate even when driven in exactly the same manner;
2. there are subtle differences in the way inspectors drive the trace on each test which causes the vehicle's emissions to vary from test-to-test; and
3. there are differences in the trace and the test equipment between the two systems being compared which result in some variation between the tests (e.g., the 2 traces work the vehicle differently).

Because vehicles' emissions vary, all I&M tests have some level of EOC, even when the same test is compared to itself. The result is that some marginal vehicles, with emissions that fluctuate near the cutpoints, will fail the test when they would have passed a reference test. Some will also pass when they would have failed a reference test, generating a false pass or "error of omission"). The majority of EOC are vehicles operating close to the cutpoints. If such vehicles were run through a number of tests using the same procedure, some of them would turn out to emit a little above the standards, on average, and some would emit a little below the standards, on average. Because it is not convenient or cost effective for motorists to perform an extended number of tests, all I&M programs accept some level of imprecision in the form of EOC in order to balance test effectiveness and motorist convenience.

Due to this normal test-to-test variation in vehicle emissions results, cutpoints are set sufficiently loose so that, even with the variation, truly clean (i.e., properly functioning vehicles) are unlikely to fail. A clean vehicle that does fail would be considered a true "false failure" because it does not require repair. In setting the cutpoints loose enough to allow for variation, I&M programs attempt to eliminate the false failure of truly clean vehicles. This means that I&M programs do not attempt to fail every vehicle with malfunctioning emissions controls, but rather aim to fail only the dirtiest of the broken vehicles. This margin of safety inherently results in the test passing some dirty vehicles.

This margin of safety means that typically EPA IM240 and MA31 cutpoints range from 3-8 times a vehicle's Federal Test Procedure (FTP) standard. For example, the MA31 NOx cutpoint for 1984-90 cars is set at approximately 3 times the FTP standard, which is the test used by EPA and auto manufacturers to certify that a vehicle's emissions controls are operating properly when it was new. In other words, the MA31 cutpoint is much *less* stringent than the FTP. This margin is set to ensure that the I&M test fails only broken, highly polluting vehicles despite the normal test-to-test variation in vehicle emissions and the differences in test and equipment types. The margin also allows for some degradation of emissions control performance due to normal wear.

6.1.3 Excess Emissions

EPA defines excess emissions as the quantity of emissions identified by the IM240 inspection that are greater than the IM240 cutpoint. An alternative test cycle, like the MA31, gets credit for identifying excess emissions if it fails the vehicle producing the excess emissions. Consider the following example:

<u>Table 12</u>
Example Excess Emission Calculation

	HC (g/mi)	CO (g/mi)	NO (g/mi)
IM240 score	0.8	23	2.7
IM240 cutpoint	1.2	20	2.5
Excess Emissions	**0.0**	**3**	**0.2**

If the MA31 test cycle failed the above vehicle, it would receive credit for identifying 3 grams per mile (g/mi) excess CO and 0.2 g/mi NOx, regardless of which pollutants the vehicle failed in the MA31 test. If the MA31 test cycle passed the above vehicle, this would be considered an error of omission in which no excess emissions are identified. For the purpose of this study, excess emissions identified by the MA31 test were calculated relative to the IM240 test with start-up cutpoints.

In granting approval of Massachusetts' I&M program SIP, EPA assigned interim MA31 test effectiveness levels of 85%, 87%, and 85% for HC, CO, and NOx, respectively, when compared to the IM240 test with start-up cutpoints. These SIP credits are the same as those established by EPA for New York's I&M program (NYTEST), which uses the same equipment as Massachusetts, but runs the IM240 drive trace. To satisfy Massachusetts' I&M program objectives, the MA31 test should meet or exceed these test effectiveness levels or "SIP target" limits.

Table 13 shows the excess emissions identified for the AZ study dataset. The excess emissions identified in g/mi for IM240 and MA31 (shown in the middle two sets of columns) are the sums of all of the HC, CO and NO excess emissions from the vehicles failing the IM240 and MA31 tests, respectively. The right-hand set of columns expresses the HC, CO and NO excess emissions identified by MA31 as percentages of those identified by IM240, thus showing the estimated "test effectiveness" of the MA31 test.

<u>Table 13</u>
Excess Emissions Identified vs. IM240 Start-Up Cutpoints
AZ Study Dataset – 612 Vehicles

Vehicle Type	Model Year	Vehicle Count	Excess Emission ID'ed by IM240 (g/mi)			Excess Emission ID'ed by MA31 (g/mi)			Excess Emissions ID'ed by MA31 (%)		
			HC	CO	NO	HC	CO	NO	HC	CO	NO
LDGV	81-84	45	3.14	12.07	7.72	1.77	4.18	6.43	56	35	83
	85-89	105	22.18	759.67	19.08	21.95	750.2	18.85	99	99	99
	90+	176	17.18	462.41	12.06	16.70	446.11	11.64	97	96	97
	All	326	42.50	1234.2	38.86	40.43	1200.5	36.92	95	97	95
LDGT1	81-84	25	4.03	272.80	0.00	3.20	264.49	0.00	79	97	-
	85-89	98	8.52	69.48	8.33	6.46	35.75	4.76	76	51	57
	90+	85	9.66	168.47	12.28	5.27	98.87	10.58	55	59	86
	All	208	22.21	510.75	20.62	14.93	399.11	15.34	67	78	74
LDGT2	81-84	4	0.00	0.00	0.00	0.00	0.00	0.00	-	-	-
	85-89	28	6.35	84.88	2.68	5.11	68.16	2.68	81	80	100
	90+	46	12.85	0.00	10.88	12.85	0.00	9.30	100	-	85
	All	78	19.20	84.88	13.56	17.96	68.16	11.97	94	80	88
Grand Total		612	83.90	1829.8	73.04	73.32	1667.7	64.24	**87**	**91**	**88**

Note: These excess emissions identified rates are for the Arizona study dataset only, which was not intended to represent the distribution of vehicles in the Massachusetts fleet. The excess emissions identified are in "grams per mile" and therefore do not account for the differences in mileage accumulation rates for different model years and vehicle types.

This analysis is included here to demonstrate how test effectiveness is calculated from IM240 and MA31 test results. However, it is not relevant to compare these total test effectiveness values to the SIP targets for the Massachusetts I&M program. The reason for this is that selection of vehicles for AZ study dataset was optimized for the conversion factor analysis; i.e. to aid in determining the most appropriate conversion factors. The

AZ study dataset was intended to contain vehicles over the *entire* range of vehicle types, vehicle model years, and vehicle emission rates that are observed in the Massachusetts program, including older, dirtier vehicles with high emissions. However, this causes a larger proportion of higher emitting vehicles in the sample to have a larger total of excess emissions that tend to be more easily identified by the test, thus overestimating the potential effectiveness of the test. In other words, a bias in this dataset towards high emitters would likely overestimate the potential effectiveness of the MA31 test. The next section of the report will look at more appropriate methods for determining the potential effectiveness of the MA31 test.

6.2 Using a 2% random sample of AZ IM240 tests

As previously mentioned, the AZ study dataset was purposely biased toward older, dirtier vehicles and was therefore not directly applicable for evaluating the effectiveness of the MA31 test. To determine the effectiveness of the test, results from the AZ study dataset would need to be applied to the specific vehicle population in the Massachusetts fleet. The best way to perform this analysis would be to take a cross section of emission results from the Massachusetts vehicle inspection program, predict IM240 scores for these data, and calculate excess emissions from those results. However, Sierra had concerns about using emission results from the Massachusetts program for this analysis because actual emission levels and failure rates, when using the correct conversion factors, appeared to be lower than expected. With the source data biased towards lower emissions, the analysis would likely underestimate the potential effectiveness of the MA31 test.

DEP is investigating reasons for the lower than expected emission scores with the Massachusetts program data. Likely reasons for this are quality assurance/quality control issues such as in-use MASS99 equipment problems, improper test delivery by inspectors, and motorist compliance, which are outside the scope of this study.

The best alternative available for determining the potential effectiveness of the MA31 test was to apply the AZ study results to an existing dataset of randomly selected vehicles properly tested in a more controlled environment. This would eliminate the bias of the AZ study dataset towards older, higher emitting vehicles.

Sierra recommended using a dataset from Arizona's I&M program evaluation that contained a random sampling of 3,734 vehicles subject to their I&M program in 1999. This sample consists of approximately 2% of the vehicles in the Phoenix, Arizona I&M program that were randomly given a full duration IM240 test. Because the vehicles were selected randomly, this 2% sample was representative of the Arizona vehicle population both in terms of vehicle population makeup (e.g. model year and type) and vehicle condition at the time of the test.

To use the AZ 2% dataset to determine the potential test effectiveness of the MA31 test, a method was developed for converting the IM240 scores (collected in AZ 2% random dataset) to MA31 scores representative of those that were collected during the AZ study. The method used for this purpose comprised two portions: 1) using the AZ study dataset of 612 vehicles to develop regressions to convert IM240 scores from the AZ 2% random sample into MA31 scores and 2) modeling the expected variability of MA31 test results using a "Monte Carlo" simulation.

6.2.1 Methodology for Estimating MA31 scores from IM240 and Modeling Variability

Sierra subcontracted with an independent statistical consultant, RW Crawford, to determine the best method for converting IM240 scores to MA31 scores and evaluate the efficacy of using a Monte Carlo simulation to model expected variation in the MA31 results.

A Monte Carlo simulation takes regression results and mimics the "random" scatter that appears in the original dataset used to create the regression. When a dataset produces a regression equation, the equation does not always predict the actual data points perfectly. The difference between a predicted data point and an actual data point is called a residual. When the regression is done properly, the magnitude of residual values should be randomly distributed across the range of data points. Thus, by characterizing the distribution of residuals, this distribution can be applied to results predicted by the regression equation to simulate the scatter present in the original data.

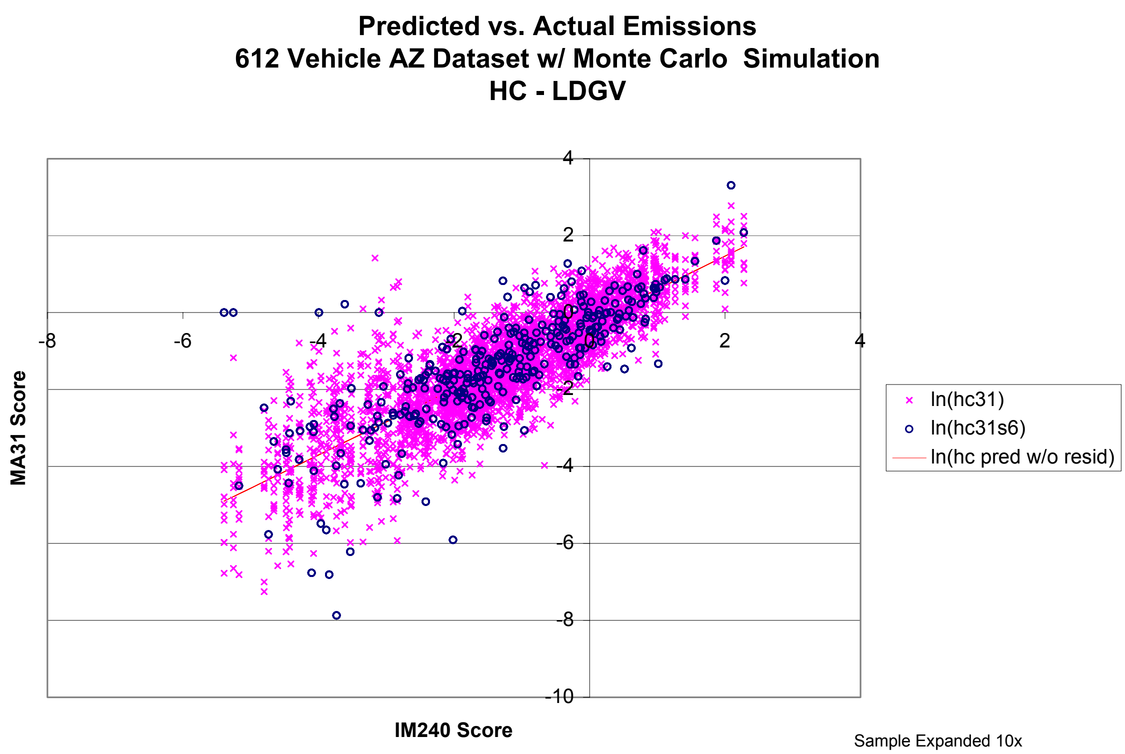
It is important to include the residuals when predicting emission scores to better reflect reality. If the residuals were ignored, predicted scores would correlate perfectly with the scores used to make the prediction. Further analyses based upon this assumption would make the MA31 test appear to be a perfect surrogate for the IM240 test, which would neither make sense nor be true.

The methodology for converting IM240 scores into MA31 scores and modeling variability using a Monte Carlo simulation is described in the report by Crawford in Appendix B. After evaluating several alternatives, Crawford selected log-log, single dependent variable regressions for the regression equations because they fit the data relatively well in terms of the overall trends in the data and the characteristics of the residuals. These equations are more complex than the linear equations used by the MASS99 equipment and in the MA31 to IM240 conversion factor analyses presented in Section 5. Separate regressions were performed for passenger cars (LDGV) and trucks (LDGT1 and LDGT2) for each of the three pollutants. Residuals were modeled separately for low, middle, and high ranges of emissions as described in the Crawford report.

Figure 9 shows one of the log-log regression results along with predicted scores using the Monte Carlo simulation for HC emissions from LDGVs in the 612-vehicle AZ study dataset. The variable "hc31s6" represents the measured MA31 HC score. The variable labeled "hc31" represents the predicted MA31 HC score, including the residual. Finally, the variable "hc pred w/o resid" represents the predicted score without the residual added. Appendix C contains all the IM240 to MA31 regression plots from the Monte Carlo simulation using the AZ study dataset.

For the analysis, the dataset was increased by a factor of 10 to even out the effects of random variation. In other words, 10 predictions were made for each original test result. For this reason, there are significantly fewer *actual* results than there are *predicted* results.

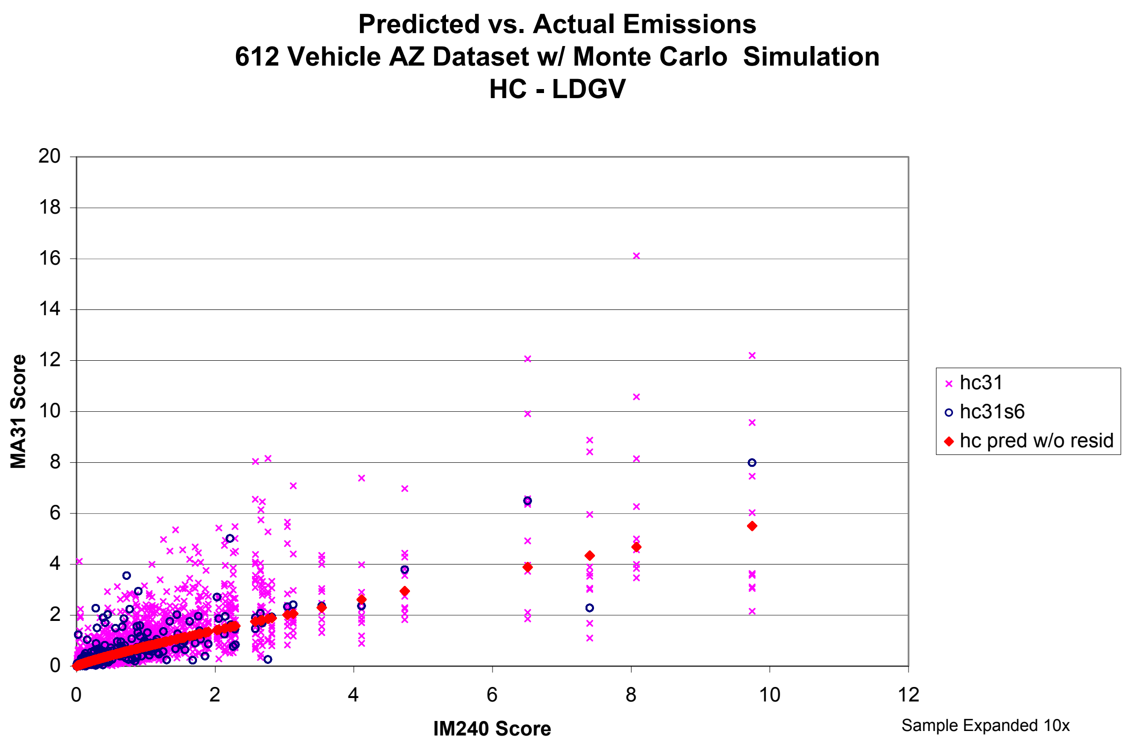
Figure 9



As can be seen above, the cloud of predicted points (the smallest points) seems reasonably well distributed throughout the actual points (the larger points), which, in turn, line up fairly well along the regression line. This validates the accuracy of the regression and the Monte Carlo simulation results.

Figure 10 details the same information after conversion back to a non-log scale.

Figure 10



After developing the regressions and performing the Monte Carlo simulation, it was necessary to take a look at how the results compared to those from the original AZ study dataset, in terms of failure rates, EOC, and excess emissions identified to evaluate the validity of the method. To perform this analysis, the IM240 to MA31 regressions developed by Crawford were used to convert IM240 results from the 612-vehicle AZ study into MA31 results. The Monte Carlo simulation was then used to add residuals to those MA31 results to simulate the variability observed in the original dataset.

Table 14 shows the MA31 failure rates generated using the predicted scores. The dataset has been expanded by a factor of 10 using the Monte Carlo simulation to model sufficient random variation.

Using the simulation to predict results, the overall MA31 failure rate was 19.1% compared to 21.2% for the original 612-sample dataset (see Table 8). The MA31 failure rate using the simulation is still very fairly close to the 20.8% failure rate for the IM240 test with start up cutpoints, the test the MA31 is designed to mimic.

Table 14
MA31 Failure Rates
Monte Carlo simulation using the AZ Study Dataset – 6,120 Samples

Vehicle Type	Model Years	Vehicle Count	MA31 Failure Rates				IM240 Failure Rates	
			HC	CO	NOx	Overall	Start-Up Cutpoints	Final Cutpoints
LDGV	81 - 84	450	9.6 %	7.1 %	12.9 %	26.4 %	28.9 %	62.2 %
	85 - 89	1,050	12.4 %	12.2 %	11.3 %	27.7 %	28.6 %	58.1 %
	90 +	1,760	6.2 %	6.8 %	8.5 %	17.0 %	15.3 %	29.6 %
	All	3,260	8.7 %	8.6 %	10.0 %	21.7 %	21.5 %	43.3 %
LDGT1	81 - 84	250	5.6 %	12.4 %	1.6 %	18.8 %	24.0 %	52.0 %
	85 - 89	980	6.8 %	4.2 %	5.3 %	15.2 %	21.4 %	51.0 %
	90 +	850	6.8 %	5.5 %	8.6 %	16.5 %	18.8 %	23.5 %
	All	2,080	6.7 %	5.7 %	6.2 %	16.2 %	20.7 %	39.9 %
LDGT2	81 - 84	40	0.0 %	2.5 %	0.0 %	2.5 %	0.0 %	0.0 %
	85 - 89	280	7.5 %	4.3 %	6.4 %	14.6 %	14.3 %	32.1 %
	90 +	460	10.0 %	0.2 %	9.3 %	17.6 %	21.7 %	30.4 %
	All	780	8.6 %	1.8 %	7.8 %	15.8 %	18.0 %	29.5 %
Grand Total		6,120	8.0 %	6.7 %	8.4%	**19.1 %**	**20.8 %**	**40.4 %**

Note: These failure rates are for the Monte Carlo simulation using Arizona study dataset. The simulation was not intended to represent the distribution of vehicles in the Massachusetts fleet. Also, since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates.

Tables 15 shows the MA31 errors of commission vs. the IM240 test using start-up and final cutpoints.

Table 15
MA31 EOC Rates
Monte Carlo simulation using the AZ Study Dataset – 6,120 Samples

Vehicle Type	Model Year	Vehicle Count	EOC Rates vs. EPA IM240 Start-Up Cutpoints				EOC Rates vs. EPA IM240 Final Cutpoints			
			HC	CO	NOx	Overall	HC	CO	NOx	Overall
LDGV	81-84	450	2.0	4.0	3.6	9.3	0.4	1.1	0.7	2.2
	85-89	1,050	2.5	2.7	3.5	8.4	0.2	0.4	0.7	1.2
	90+	1,760	1.6	2.6	3.0	6.6	0.5	1.3	1.3	2.8
	Total	3,260	1.9	2.8	3.3	7.6	0.4	1.0	1.0	2.2
LDGT1	81-84	250	4.0	2.8	1.6	8.0	2.0	1.6	0.4	3.6
	85-89	980	2.9	1.3	1.9	5.6	0.7	0.1	0.2	1.0
	90+	850	1.2	0.5	2.5	4.1	0.5	0.1	1.9	2.5
	Total	2,080	2.3	1.2	2.1	5.3	0.8	0.3	0.9	1.9
LDGT2	81-84	40	0.0	2.5	0.0	2.5	0.0	2.5	0.0	2.5
	85-89	280	1.4	0.4	3.2	5.0	0.0	0.0	1.4	1.4
	90+	460	2.2	0.2	2.6	5.0	1.1	0.2	1.7	3.0
	Total	780	1.8	0.4	2.7	4.9	0.6	0.3	1.5	2.4
Grand Total		6,120	2.0	1.9	2.8	**6.5**	0.5	0.7	1.0	**2.1**

Note: These EOC rates are for the Monte Carlo simulation using Arizona study dataset. The simulation was not intended to represent the distribution of vehicles in the Massachusetts fleet.

As Table 15 shows, the predicted MA31 EOC rates were 6.5% and 2.1% when compared to the IM240 test with start up and final cutpoints, respectively. This matches reasonably well with 7.0% and 2.6% calculated from the original dataset (see Table 11).

Finally, Table 16 shows the predicted excess emission identified by the MA31 test compared to IM240 with start-up cutpoints.

Table 16

Excess Emissions Identified vs. IM240 Start-Up Cutpoints

Monte Carlo simulation using the AZ Study Dataset – 6,120 Samples

Vehicle Type	Model Year	Vehicle Count	Excess Emission ID'ed by IM240 (g/mi)			Excess Emission ID'ed by MA31 (g/mi)			Excess Emissions ID'ed by MA31 (%)		
			HC	CO	NOx	HC	CO	NOx	HC	CO	NOx
LDGV	81-84	450	31.39	120.7	77.25	22.48	96.56	48.63	72	80	63
	85-89	1,050	221.8	7577	190.8	203.2	7046	146.6	92	93	77
	90+	1,760	171.7	4624	120.6	169.3	4466	84.94	99	97	70
	All	3,260	425.0	12340	388.6	395.0	11610	280.2	93	94	72
LDGT1	81-84	250	40.27	2727	0.00	33.69	2311	0.00	84	85	-
	85-89	980	85.20	694.8	83.35	42.67	482.9	46.16	50	69	55
	90+	850	96.93	1685	122.8	81.76	1417	98.52	85	84	80
	All	2,080	222.1	5107	206.2	158.1	4211	144.7	71	82	70
LDGT2	81-84	40	0.00	0.00	0.00	0.00	0.00	0.00	-	-	-
	85-89	280	63.48	848.8	26.79	57.73	782.0	15.03	91	92	56
	90+	460	114.5	0.00	93.84	103.4	0.00	57.74	90	0	62
	All	780	178.0	848.8	120.6	161.1	782.0	72.77	91	92	60
Grand Total		6,120	825.0	18298	715.4	714.3	16603	497.7	**87**	**91**	**70**

Note: These excess emissions identified rates are for the Monte Carlo simulation using Arizona study dataset. The simulation was not intended to represent the distribution of vehicles in the Massachusetts fleet. The excess emissions identified are in "grams per mile" and therefore do not account for the differences in mileage accumulation rates for different model years.

Predicted results for HC and CO using the Monte Carlo simulation matched the results obtained using the original dataset. However, the predicted overall NOx results were noticeably lower than the original dataset (70% vs. 88%). This occurred because a small number of vehicles in the AZ study dataset had large quantities of excess NOx emissions that showed statistically anomalous results when considering the behavior of other vehicles in the dataset. The Monte Carlo simulation corrected this statistical anomaly and thus lowered the excess emission identification rate for NOx to a rate that might have been expected had the original dataset been larger. Overall, the Monte Carlo simulation methodology for converting IM240 scores from MA31 scores and applying expected variability appears to be valid with respect to failure rates, EOC rates, and excess emissions identified when tested against the AZ study dataset of 612 MA31 and IM240 matched tests.

6.2.2 Monte Carlo Simulation applied to the AZ 2% Sample Dataset

The next portion of the analysis consisted of applying the Monte Carlo simulation to the AZ 2% random sample dataset to estimate the effectiveness of the MA31 test. As mentioned previously, this dataset contains vehicles that were randomly selected rather than being biased toward older, high emitting vehicles and, therefore, should better represent the Massachusetts fleet.

To perform the test effectiveness analyses using the AZ 2% sample dataset, the Monte Carlo simulation was used to predict raw MA31 scores from the IM240 scores in the dataset. Each of the 3,734 samples in the dataset was used to predict 10 separate raw MA31 scores, creating 37,340 predicted raw MA31 scores. The MA31 to IM240 conversion factors developed in Section 5.2 were then used to calculate equivalent IM240 scores (i.e. converted MA31 scores) from the raw MA31 scores, the same way the MASS99 test system converts emissions scores during each actual test. These 37,340 MA31 samples were then used to determine failures rates, EOC rates, and excess emissions identified.

Table 17 shows the predicted MA31 failure rates from the Arizona 2% random sample data.

<u>Table 17</u>

MA31 Failure Rates

Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

Vehicle Type	Model Years	Vehicle Count	MA31 Failure Rates				IM240 Failure Rates	
			HC	CO	NOx	Overall	Start-Up	Final
LDGV	81 – 84	2,000	24.6 %	18.8 %	19.7 %	47.5 %	44.5 %	81.5 %
	85 – 89	8,240	13.2 %	12.3 %	16.1 %	32.4 %	27.8 %	57.0 %
	90 +	14,300	6.5 %	5.8 %	7.8 %	16.5 %	9.0 %	20.9 %
	All	24,540	10.2 %	9.0 %	11.5 %	24.4 %	18.2 %	38.0 %
LDGT1	81 – 84	640	10.5 %	19.5 %	6.6 %	30.5 %	29.7 %	65.5 %
	85 – 89	2,880	10.0 %	6.6 %	11.4 %	24.1 %	23.6 %	53.5 %
	90 +	5,960	3.9 %	1.3 %	8.8 %	12.6 %	9.7 %	19.1 %
	All	9,480	6.2 %	4.1 %	9.4 %	17.3 %	15.3 %	32.7 %
LDGT2	81 – 84	500	6.8 %	15.8 %	8.4 %	28.2 %	26.0 %	78.0 %
	85 – 89	960	11.6 %	9.3 %	8.0 %	24.4 %	20.8 %	54.2 %
	90 +	1,860	8.3 %	2.6 %	6.4 %	14.6 %	8.6 %	18.3 %
	All	3,320	9.0 %	6.5 %	7.2 %	19.5 %	14.8 %	37.7 %
Grand Total		37,340	9.1 %	7.5 %	10.6 %	**22.1 %**	**17.1 %**	**36.6 %**
Note: These failure rates are not expected to match those for the current Massachusetts test. They are based on one-chance-to-pass the MA31 test and do not include any vehicles newer than model year 2000. Since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates.								

Table 18 shows the predicted MA31 EOC rates vs. IM240 start up and final cutpoints.

Table 18
MA31 EOC Rates
Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

Vehicle Type	Model Year	Vehicle Count	EOC Rates vs. EPA IM240 Start-Up Cutpoints				EOC Rates vs. EPA IM240 Final Cutpoints			
			HC	CO	NOx	Overall	HC	CO	NOx	Overall
LDGV	81-84	2,000	4.6	3.9	5.6	12.9	0.1	0.6	0.2	0.8
	85-89	8,240	3.6	3.9	5.2	11.7	0.4	0.9	1.1	2.3
	90+	14,300	3.3	3.5	4.1	10.1	1.6	2.1	2.1	5.6
	Total	24,540	3.5	3.7	4.6	10.8	1.1	1.6	1.6	4.1
LDGT1	81-84	640	1.7	4.5	3.9	9.7	0.2	0.6	0.3	1.1
	85-89	2,880	3.3	1.9	3.9	8.6	0.8	0.2	0.7	1.7
	90+	5,960	2.3	0.7	4.0	6.6	1.4	0.2	1.9	3.5
	Total	9,480	2.5	1.3	4.0	7.4	1.1	0.3	1.5	2.8
LDGT2	81-84	500	3.2	9.2	3.2	14.8	0.2	2.2	0.4	2.8
	85-89	960	5.9	3.2	3.4	11.4	1.8	0.1	0.8	2.7
	90+	1,860	4.7	0.6	3.9	8.8	3.1	0.3	2.6	5.6
	Total	3,320	4.8	2.7	3.6	10.5	2.3	0.5	1.7	4.4
Grand Total		37,340	3.4	3.0	4.4	**9.9**	1.2	1.2	1.6	**3.8**

Note: These EOC rates are not expected to match those for the current Massachusetts test. They are based on one-chance-to-pass the MA31 test and do not include any vehicles newer than model year 2000.

These data show that the predicted MA31 failure rates and EOC rates for the AZ 2% random dataset are somewhat higher than the AZ study dataset. Since the AZ 2% dataset was not designed to be biased toward dirty vehicles, there should be a larger portion of vehicles operating near the cutpoint and therefore more likelihood that they will be falsely failed.

Table 19 shows the predicted excess emissions identified for the Arizona 2% random sample. The excess emission are reported as “grams” instead of “grams per mile” to reflect mileage accumulation weightings for the specified model year ranges and vehicle types. These weightings are necessary because older vehicles (that tend to provide most of the excess emissions) are not driven as much as newer vehicles. These mileage accumulation weightings were developed by EPA for their MOBILE6 emissions factor model and are presented in Appendix D. Excess emissions as “grams” are calculated by multiplying the excess emissions as “grams per mile” by the annual mileage accumulation rate shown in Appendix D.

Table 19

Excess Emissions Identified vs. IM240 Start-Up Cutpoints

Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

Vehicle Type	Model Year	Vehicle Count	Excess Emission ID'ed by IM240 (grams * 10^5)			Excess Emission ID'ed by MA31 (grams * 10^5)			Excess Emissions ID'ed by MA31 (%)		
			HC	CO	NOx	HC	CO	NOx	HC	CO	NOx
LDGV	81-84	2,000	722	12,650	404	686	11,930	353	95	94	87
	85-89	8,240	1,470	30,010	1,380	1,350	27,710	1,180	92	92	86
	90+	14,300	765	10,980	704	726	10,250	551	95	93	78
	All	24,540	2,957	53,640	2,488	2,762	49,890	2,084	94	93	84
LDGT1	81-84	640	163	3,601	61	139	3,330	36	85	92	59
	85-89	2,880	727	5,372	753	628	5,090	605	86	95	80
	90+	5,960	298	2,381	790	269	2,369	581	90	100	73
	All	9,480	1,188	11,354	1,604	1,036	10,789	1,222	87	95	76
LDGT2	81-84	500	70	1,052	83	37	913	46	53	87	55
	85-89	960	140	1,838	189	96	1,565	154	68	85	81
	90+	1,860	359	3,179	152	343	3,091	107	96	97	71
	All	3,320	569	6,069	424	476	5,569	307	84	92	72
Grand Total		37,340	4,714	71,063	4,516	4,274	66,248	3,612	**91**	**93**	**80**

Note: These excess emissions identified rates are not expected to match those for the current Massachusetts test. They are based on one-chance-to-pass the MA31 test and do not include any vehicles newer than model year 2000. The Section 7.0 evaluates the current MA31 test design.

7.0 EVALUATION OF CURRENT MA31 TEST DESIGN

The previous analyses assumed a MA31 test sequence consisting of six traces with only one chance to pass. This approach was expected to yield the best excess emissions identification performance achievable within the current program design. However, from May 2001 until March 14, 2003, the MA31 test sequence was "Fast Pass" where the vehicle would "pass out" of the test on any of the six traces. As soon as a vehicle passed one of the MA31 traces, it passed the test and the inspection was completed. A single MA31 trace was used to pre-condition the vehicle prior to the test.

"Fast Pass" was introduced into the Massachusetts I&M program to increase motorist and inspector convenience and vehicle throughput by requiring less driving time on the dynamometer for cleaner vehicles to pass the test. Use of a "Fast Pass" test sequence was justified by the fact that most vehicles are clean and therefore should not be required to complete a longer test sequence needed to identify marginal or dirty vehicles. Many states have implemented a "Fast Pass" test sequence in their I&M programs for this reason.

Sierra used the following methodology and assumptions for the analysis of the current program:

- Interim MA31 to IM240 conversion factors developed from the full AZ study dataset regressions are in use (Section 5.1). These are the conversion factors currently being used in the Massachusetts I&M program.
- Vehicles are fully warmed up prior to starting first MA31 trace.
- Vehicles pass the MA31 test by passing any one of the six MA31 traces.
- Monte Carlo simulation of the AZ% 2% random sample dataset to generate MA31 scores for the first trace that are representative of the Massachusetts fleet (as described in the previous section).
- Subsequent MA31 trace scores developed based upon test-to-test variability as measured between MA31 traces 5 and 6 of the 612-vehicle AZ study dataset.

Raw MA31 scores from the first MA31 trace were predicted from the AZ 2% sample dataset using the Monte Carlo simulation technique. Raw MA31 scores were then converted to final MA31 scores (i.e. equivalent IM240 scores) using the MA31 to IM240 conversion factors. Final MA31 scores were then compared to the Massachusetts program cutpoints. Vehicles that passed the test on the first trace were set aside. Vehicles that failed the first trace had a second MA31 trace analyzed. The second MA31 trace scores were predicted using the first MA31 score as the baseline, and applying random variation to the scores based on the variation that occurred between traces 5 and 6 of the AZ 612-vehicle dataset. Vehicles that passed the second MA31 trace were set aside. Vehicles that failed the second MA31 trace had a third MA31 trace analyzed. This process was continued for a total of six MA31 traces, until all of the vehicles in the AZ 2% sample dataset had completed the complete MA31 test sequence. The failure rates, EOC rates, and excess emissions identified were then calculated based on these results.

Table 20 shows the excess emissions identified, failure rates, and EOC rates for the MA31 "Fast Pass" test sequence. The data show that the "Fast Pass" test sequence meets the SIP targets for HC and CO, but falls well short for NOx. Based on these data, changes are needed to the Massachusetts test design to meet the NOx SIP target.

Table 20

MA31 Test with 6 Chances to Pass ("Fast Pass") and Interim Conversion Factors

Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

	HC	CO	NOx	Overall
MA31 to IM240 Conversion Factors	0.98	0.57	0.56	-
Excess Emissions Identified	87 %	90 %	69 %	-
SIP Targets	85 %	87 %	85 %	-
Failure Rates*	9.3 %	6.6 %	7.2 %	18.0 %
EOC Rates vs. IM240 Start-Up Cutpoints	3.4 %	2.3 %	2.5 %	7.3 %
EOC Rates vs. IM240 Final Cutpoints	1.2 %	0.9 %	0.8 %	2.7 %

* Note: Since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates.

Predicted EOC rates from the AZ 2% random sample are 7.3% and 2.7% compared to IM240 start-up and final cutpoints, respectively. These are well within the generally accepted range for I&M programs. As noted previously, it is desirable to keep the overall EOC rate under 5% relative to IM240 *final* cutpoints. It is acceptable to have an overall EOC rate somewhat greater than 5% relative to IM240 *start-up* cutpoints since these vehicles are likely to have emission control malfunctions that can be repaired.

8.0 INVESTIGATION OF ALTERNATIVES TO IMPROVE TEST EFFECTIVENESS

In considering alternatives to improve test effectiveness, over a dozen options were developed and analyzed. The following sections describe and evaluate four test design changes considered most feasible to improve test effectiveness and meet SIP targets.

8.1 Increase Effectiveness of the MA31 Test

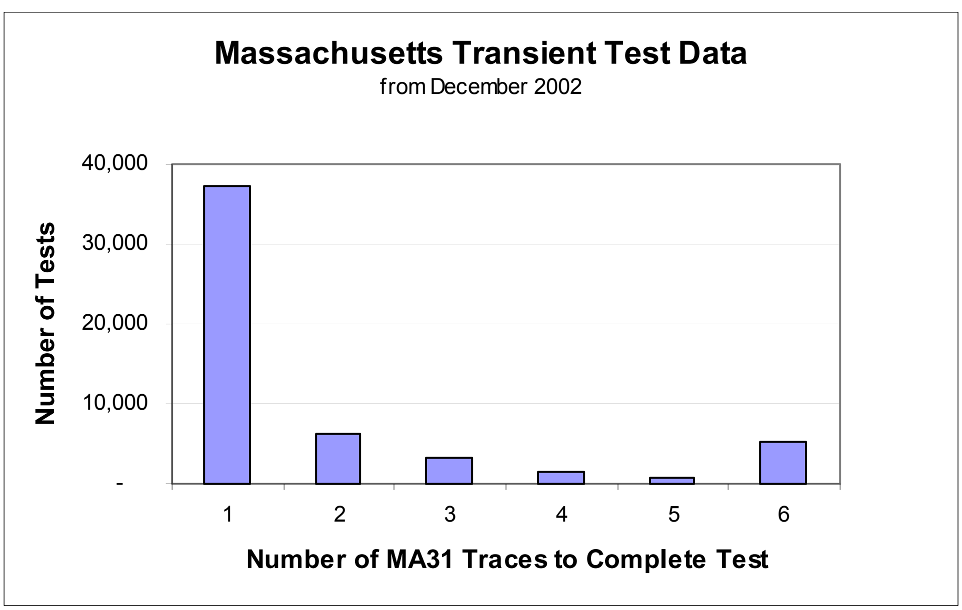
The first logical step was to look at changes to the existing MA31 test sequence that would increase test effectiveness to meet SIP targets. Because these changes are within the existing structure of the MA31 test sequence, they would be the quick and easy to implement.

8.1.1 Elimination of "Fast Pass"

This analysis studied the effect of switching from "Fast Pass" back to the original test sequence that allows only two chances to pass.

Massachusetts' "Fast Pass" test sequence requires an initial pre-conditioning MA31 trace to warm up the vehicle followed by up to 6 chances to pass the MA31 trace. As soon as the vehicle passes a trace, the test is completed and the vehicle passes the inspection. Figure 11 shows that, with "Fast Pass", the large majority of vehicles pass the test after the first MA31 trace, saving time for the motorist and inspector. Vehicles failing the emissions test must fail all six MA31 traces.

Figure 11
MA31 Trace-by-Trace Analysis
For "Fast Pass" – Six Chances to Pass



An unintended side effect of the "Fast Pass" test sequence is that some dirtier vehicles hovering around the cutpoints end up passing the test. This occurs because there is inherent trace-to-trace variability with vehicle emissions, especially for higher emitting vehicles. By giving these vehicles six chances to pass, there is a greater chance that one of the traces will fall below the cutpoints and pass the test.

To investigate the magnitude of this effect, "Fast Pass" was shut off for a period of two weeks in February 2002. The test sequence replacing "Fast Pass" during this period allowed only two chances to pass. This sequence required the vehicle to first complete two pre-conditioning MA31 traces. The third trace was then used as the first official test trace. If the vehicle failed, it was required to complete two post-conditioning traces and was given a second chance to pass the inspection on the sixth and final trace. Under this scenario, all vehicles completed three traces and the dirtier vehicles completed six traces. This was the same test sequence that was used before "Fast Pass" was implemented. The advantage of a two-chances-to-pass sequence over one-chance-to-pass is that clean vehicles (i.e. the majority of the fleet) do not have to complete all six traces to pass.

Table 21 shows a breakdown of the 2,044 vehicles that failed the MA31 test sequence during the two-week period when "Fast Pass" was turned off. The data show that 1,744 (86.2%) of the *failing* vehicles failed all six MA31 traces during the test. These are vehicles that would have failed "Fast Pass" since they failed all six chances to pass. However, 280 (13.8%) of the failing vehicles failed fewer than six MA31 traces (i.e. passed at least one of the traces) and therefore would have passed under "Fast Pass." The breakdown of this 13.8% shows the large majority failed five of the six traces, indicating most of the 280 vehicles are fairly consistent high emitters. Allowing these vehicles to pass the test has a notable effect on the transient test failure rate and effectiveness at identifying excess emissions. The failure rate increased from 5.0% to 6.7 % during the period when "Fast Pass" was turned off.

Table 21

Two-week period when "Fast Pass" was replaced with Two-Chances-To-Pass
Breakdown of Failing Vehicles

	Number of Vehicles	Percent of Total Failures
Failed all six traces	1,744	**86.2%**
Failed 2 to 5 traces	280	**13.8%**
Failed 5 of 6 traces	184	9.1%
Failed 4 of 6 traces	69	3.4%
Failed 3 of 6 traces	24	1.2%
Failed 2 of 6 traces	2	0.1%

To determine the effect that eliminating "Fast Pass" has on the overall test effectiveness, it was necessary to again use the Monte Carlo simulation of the AZ 2% dataset to predict MA31 results. This analysis followed the same procedure used to predict scores for six chances to pass (described in Section 7.0), except the process was stopped after the first two traces, since the replacement test sequence allows only two chances to pass.

Table 22 presents the excess emissions identified, failure rate, and EOC for a MA31 test sequence that allows only two chances to pass. The data show that this scenario increases the excess emissions identified for all three pollutants when compared to "Fast Pass," but it still doesn't meet the SIP target for NOx.

Table 22

MA31 Test with Two-Chances-to-Pass and Interim Conversion Factors

Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

	HC	CO	NOx	Overall
MA31 to IM240 Conversion Factors	0.98	0.57	0.56	-
Excess Emissions Identified	91 %	93 %	75 %	-
SIP Targets	85 %	87 %	85 %	-
Failure Rates *	10.3 %	7.5 %	8.5 %	20.9 %
EOC Rates vs. IM240 Start-Up Cutpoints	4.0 %	3.0 %	3.2 %	9.2 %
EOC Rates vs. IM240 Final Cutpoints	1.4 %	1.1 %	1.1 %	3.5 %
* Note: Since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates.				

The two-chances-to-pass sequence using interim conversion factors increases the predicted failure rate from 17.5% to 20.5% when compared to "Fast Pass". This is consistent with the results observed during the two-week period when "Fast Pass" was turned off. The predicted EOC rates increase slightly when compared to "Fast Pass," but are still within acceptable limits.

Table 23 present the two-chances-to-pass sequence using final conversion factors that were developed from the full AZ study dataset of 612 vehicles tested (section 5.2). This analysis is presented because it is anticipated that final conversion factors will be implemented following the release of this report.

Table 23

MA31 Test with Two-Chances-to-Pass and Final Conversion Factors

Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

	HC	CO	NOx	Overall
MA31 to IM240 Conversion Factors	0.87	0.53	0.60	-
Excess Emissions Identified	88 %	91 %	78 %	-
SIP Targets	85 %	87 %	85 %	-
Failure Rates *	8.5 %	6.8 %	10.1 %	20.5 %
EOC Rates vs. IM240 Start-Up Cutpoints	3.0 %	2.5 %	4.1 %	8.9 %
EOC Rates vs. IM240 Final Cutpoints	1.0 %	1.0 %	1.5 %	3.3 %
* Note: Since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates.				

Compared to the interim conversion factors, the final conversion factors for HC and CO are slightly lower and slightly higher for NOx. This affected the excess emissions identified in the same manner. With the final conversion factors, the excess emissions identified for HC and Co dropped slightly but still meet the SIP targets. The excess emissions identified for NOx increased slightly to 78%, but still falls short of the 85% target. Therefore, additional program changes would be required to meet the NOx SIP target.

8.1.2 Adjust MA31 Cutpoints to Meet SIP Targets

Another strategy that was investigated is to lower the MA31 cutpoints to meet the SIP targets for excess emissions identified. Lowering the cutpoints would cause more vehicles to fail the MA31 test and, thus, increase the total excess emissions identified.

An indirect approach was taken for this analysis. Because there are 13 different cutpoint categories in the Massachusetts program based on the vehicle model year and type, this analysis was more simply performed by keeping the existing cutpoints and raising the conversion factors to meet the SIP targets. Raising the conversion factors (and thus emissions) and keeping the cutpoints the same is essentially equivalent to lowering the cutpoints and keeping the conversion factors the same.

In addition to adjusting the conversion factors, this analysis assumes only one chance to pass the MA31 sequence, which is the most stringent form of the test. In other words, all vehicles would have to drive six MA31 traces to complete the inspection. The methodology used to explore alternative conversion factors was to iteratively run the simulation previously developed while increasing the conversion factors until the SIP excess emission targets were reached.

Table 24 presents the excess emissions identified, failure rate, and EOC rates for this scenario. To meet the SIP target of 85% excess NOx emissions identified, the MA31 to IM240 conversion factor for NOx was increased from 0.60 to 0.69 for this analysis. The conversion factors for HC and NOx were not adjusted because they already met SIP targets with a one-chance-to-pass test sequence.

Table 24

Test Alternative: Adjust MA31 Cutpoints to Meet SIP Targets

Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

	HC	CO	NOx	Overall
MA31 to IM240 Conversion Factors	0.87	0.53	0.69	-
Excess Emissions Identified	91 %	93 %	87 %	-
SIP Targets	85 %	87 %	85 %	-
Failure Rates *	9.1 %	7.5 %	14.6 %	25.3 %
EOC Rates vs. IM240 Start-Up Cutpoints	3.4 %	3.0 %	6.9 %	12.2 %
EOC Rates vs. IM240 Final Cutpoints	1.2 %	1.2 %	2.8 %	4.9 %

* Note: Since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates that would occur if this test design alternative were implemented.

Adjusting the NOx conversion factor had a noticeable effect on the NOx failure rate and overall failure rate, increasing them by approximately 4 percentage points when compared to the MA31 one-chance-to-pass analysis performed in Section 6.2.2 (Table 17). Increasing the NOx conversion factor also had a significant effect on the NOx and overall EOC rates. The NOx EOC rate increased to more than double the EOC rates for HC and CO and increased the overall EOC rate vs. IM240 final cutpoints to nearly 5%.

The main advantage of lowering the NOx cutpoint to meet SIP targets is that it can be easily implemented without changes to the software, equipment, or test procedure. However, several disadvantages are:

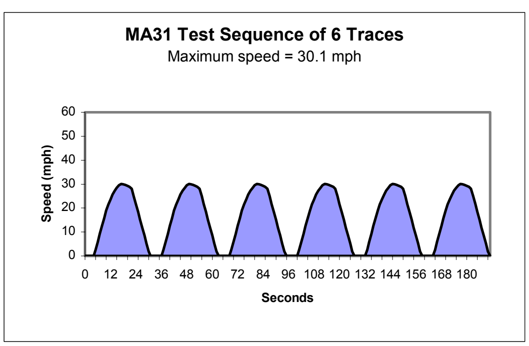
- this increases the overall EOC rate to nearly 5% (relative to IM240 final cutpoints) which is not considered acceptable for an I&M program and
- this would require a change to the Massachusetts I&M program regulation.

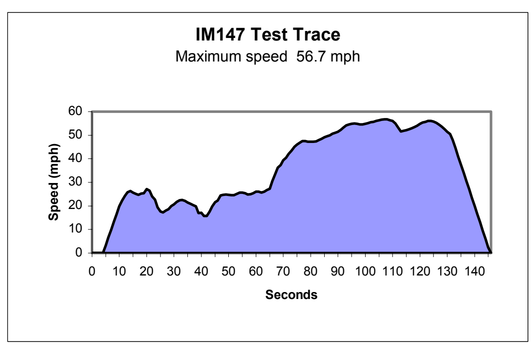
8.2 Replace the MA31 test with MA147

The IM240 trace is more effective than the MA31 trace at identifying excess NOx emissions because it exercises the vehicle at high speed and load similarly to what occurs on the FTP. The IM147 trace is simply the last 147 seconds of the IM240 trace which represents the second phase or high speed portion of the test. For this reason, IM147 results can be directly calculated from IM240 results. The advantages of the IM147 trace over the IM240 trace in an I&M program are that it decreases test time while maintaining its effectiveness at identifying excess emissions. For this reason, Arizona has switched from the IM240 trace to IM147 for all of its regular enhanced vehicle inspections.

This analysis studied the effect of replacing the MA31 test sequence with a single IM147 trace on the MASS99 equipment (i.e. MA147 test). Figure 12 compares the MA31 traces to the IM147 trace.

Figure 12
Comparison of MA31 and IM147 Tests





To perform this analysis, it was first necessary to use the AZ study dataset to develop conversion factors between the IM147 trace using MASS99 equipment (MA147) and the IM240 trace using IM240 equipment. For the conversion factor regressions, a threshold was set to include only data points that were less than or equal to 1.5 times the maximum IM240 cutpoints. As with the earlier conversion factor analyses, the rationale for setting this limit was to prevent a small number of high emission results from skewing the regression. Appendix E contains the regressions for the three pollutants and resulting conversion factors and correlation coefficients.

To determine the failure rate, EOC rate, and excess emissions identified, the same general procedure was used as for the MA31 analyses in the previous section. In this case, however, it was necessary to create a separate Monte Carlo simulation to predict MA147 scores from IM240 scores. The log-log plots of MA147 vs. IM240 data for the Monte Carlo simulation are presented in Appendix F.

Once this was done, the simulation was run on the data and analyzed the same way as was done with the MA31 results.

Table 25 presents the conversion factors, excess emissions identified, failure rates and EOC rates for the MA147 trace. The data show this scenario comfortably exceeds SIP targets for all three pollutants. It achieves this level of test effectiveness with lower failure rates and EOC rates than the MA31 test.

Table 25

Test Alternative: Replacing the MA31 test with MA147

Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

	HC	CO	NOx	Overall
MA147 to IM240 Conversion Factors	1.67	0.66	0.80	-
Excess Emissions Identified	92 %	93 %	89 %	-
SIP Targets	85 %	87 %	85 %	-
Failure Rates *	7.7 %	5.7 %	10.6 %	19.4 %
EOC Rates vs. IM240 Start-Up Cutpoints	2.2 %	1.5 %	3.3 %	6.6 %
EOC Rates vs. IM240 Final Cutpoints	0.6 %	0.5 %	0.8 %	1.8 %
* Note: Since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates that would occur if this test design alternative were implemented.				

The main advantages of replacing the MA31 test with MA147 are that the SIP targets can be comfortably met with lower failure rates and EOC rates. The disadvantages with the MA147 test are:

- **Safety:** The MA147 test operates at almost twice the speed as the MA31 trace, requiring the vehicle to be properly restrained. Although tiedown straps are supplied to shops in the Massachusetts program, they are recommended but not required for the MA31 test. Inspectors will have to be retrained to always use tiedown straps for the MA147 test. This issue is not insurmountable considering New York uses an IM240 test in its network of decentralized stations.

- **Equipment Durability:** The high-speed operation of the MA147 trace is an additional strain on the dyne and VMAS that may lead to more frequent repairs, increase equipment downtime for the network, and decrease equipment life.

- **Noise:** Due to the high-speed operation of the MA147 trace, considerably more noise is generated by the vehicle and dyne. In most shop environments, OSHA's 8-hour exposure limit of 85 decibels will be exceeded, requiring the inspector and other shop personnel to wear hearing protection to avoid long-term hearing loss. Some motorists, if they are allowed to observe testing, may also become concerned about their vehicle due to the increase in noise from the test.

In addition, DEP will need to investigate changes to the dyne calibration procedure to ensure accurate loading at the higher test speeds and changes to VMAS test limits to account for the higher exhaust flows and temperatures encountered during the MA147 test.

8.3 Implement MA31 with MA147 Second Chance Test

This analysis studied the effect of a hybrid solution that uses the MA31 test in combination with the MA147 test to more efficiently identify polluting vehicles. With this scenario, the MA31 test is used as a clean screen. If the vehicle passes the MA31 test, then the vehicle passes its emissions inspection. If the vehicle fails the MA31 test, it is given a second chance to pass the emissions inspection using the MA147 trace. There are several advantages to this testing scenario. The MA31 test can quickly pass the clean vehicles, which represents the majority of the fleet. The second-chance MA147 test, which is better at identifying excess NOx emissions than MA31, is reserved only for those marginal vehicles that did not pass the MA31 test. This scenario uses the longer, more effective test only for the vehicles that need it, which reduces some of the practical disadvantages with the MA147 test discussed in the previous section.

For this analysis, the MA31 test sequence assumes one-chance-to-pass and the MA147 test operates at the same stringency as IM240 with start-up cutpoints. The stringency of the MA31 test, however, is increased to cause 25% of the vehicles to fail the MA31 test and require the second-chance MA147 test. In other words, it will pass the cleanest 75% of the vehicles with the MA31 test only. The 25% overall failure rate for the MA31 is achieved by adjusting the MA31 conversion factors. Since all vehicles failing the MA31 portion of the test are then evaluated on the MA147 test, no EOC occur on the MA31 portion of the test. The MA147 to IM240 conversion factors are the same as those developed for the previous analysis (see Section 8.2).

Table 26 presents the conversion factors, excess emissions identified, failure rates, and EOC rates for this scenario.

Table 26

Test Alternative: MA31 Clean Screen with MA147 Second Chance test

Monte Carlo simulation using the AZ 2% Sample Dataset – 37,340 Samples

	HC	CO	NOx	Overall
MA31 to IM240 Conversion Factors	0.80	0.38	0.74	-
MA147 to IM240 Conversion Factors	1.67	0.66	0.80	-
Excess Emissions Identified	88 %	88 %	87 %	-
SIP Targets	85 %	87 %	85 %	-
Failure Rates	7.3 %	5.1 %	10.5 %	18.4 %
EOC Rates vs. IM240 Start-Up Cutpoints	2.1 %	1.3 %	3.3 %	6.3 %
EOC Rates vs. IM240 Final Cutpoints	0.6 %	0.4 %	0.8 %	1.8 %

* Note: Since this study analyzed only the drive trace and equipment under controlled circumstances, these figures were not designed to, nor do they, reflect the program's actual failure rates that would occur if this test design alternative were implemented.

The data show that the excess emission identified for all pollutants exceed the SIP targets. The main advantage of this scenario is that, like the straight MA147 test, SIP targets are met with lower failure rates and EOC rates. Also, this is achieved with only 25% of the fleet having to receive the MA147 test. The main disadvantages with this scenario are the same as for the straight MA147 test discussed in the previous section, though to a lesser degree since the MA147 test would only be administered for 25% of the vehicles tested. In addition, this option would require software changes to the analyzer and network database to allow for two sets of conversion factors (MA31 test and IM147 test).

9.0 CONCLUSIONS

This study provided data necessary to: 1) develop specific software conversion factors between the MA31 test and the IM240 test, 2) perform an evaluation of the current Massachusetts MA31 test to determine the test effectiveness in terms of ability to identify excess emissions, and 3) investigate alternatives to improve the effectiveness of the MA31 test.

9.1 MA31 to IM240 Conversion Factors

Based on an analysis of a partial dataset collected before testing was completed, “interim” conversion factors were developed and implemented in July 2001. These interim conversion factors were significantly lower than the initial conversion factors used in the Massachusetts program. This change improved the accuracy of the MA31 test, but also had the effect of lowering emission scores and the failure rate for the MA31 test. The final analysis of the full AZ study dataset (612 samples) yielded conversion factors that were only slightly different than the “interim” factors. The final conversion factors are scheduled to be implemented in Summer 2003.

9.2 Evaluation of Current I&M Test

To evaluate the effectiveness of the MA31 test in the Massachusetts I&M program, it was necessary to use an unbiased random sampling of emission tests (the AZ 2% sample dataset) and develop a methodology to to predict MA31 scores from IM240 scores. A “Monte Carlo” simulation was used to predict realistic MA31 scores from this dataset that contained the same variability or “scatter” in emission results that was observed with the 612 vehicles tested in the study. Results from this analysis showed that the MA31 test with “Fast Pass” did not meet test effectiveness targets defined in the SIP for HC and NOx.

9.3 Investigation of Alternative to Improve Test Effectiveness

Several program changes were explored to determine the optimal method for meeting the SIP targets. By eliminating “Fast Pass” and changing the MA31 test to a two-chances-to-pass sequence, the excess HC emissions identified increased so that it met the SIP target. However, the excess NOx emissions identified still were below the SIP target. As an interim program improvement, DEP replaced “Fast Pass” with a two-chances-to-pass sequence for the MA31 test on March 14, 2003.

The report investigated three additional options to meet the SIP target for NOx:

1. lower the MA31 NOx pass/fail cutpoint to fail more vehicles,

2. replace the MA31 test with MA147, a shortened version of the IM240, and

3. use the MA31 test as a clean screen and the MA147 as a second chance test.

Each of these options has advantages and disadvantages which DEP plans to explore during 2003.

10.0 STUDY LIMITATIONS

As is typical with many studies, there were limitations with the data collected and methods used for analyses in this study that are described below.

10.1 AZ Study Dataset

The study was designed to test up to 1,000 vehicles covering a wide range of vehicle types, model years, and emission rates. However, due to unanticipated repairs needed with the MASS99 equipment, difficulty in recruiting high emitting vehicles, and a commitment by Gordon-Darby to use the test lane for another project starting in August 2001, only approximately 850 vehicles were tested in the study. Additional problems with random occurrences of data file corruption on the MASS99 system and technician errors entering the 17 digit Vehicle Identification Numbers (VINs) consistently for both test systems reduced the total number of valid matched tests to 612. This dataset of 612 samples did not have an equal distribution of low, medium, and high emitting vehicles due to the difficulty of recruiting high emitting vehicles.

The lower proportion of high emitters in the dataset than planned may have affected the conversion factor analyses. Ideally, the regressions used to determine the conversion factors would have an equal distribution of data points throughout the range from low to high emissions. The regression equations tend to be less accurate where there are fewer data points.

The reduced number of total samples in the dataset and the lower proportion of high emitters also likely affected the test effectiveness analyses. The reduced number of total samples caused several of the vehicle type and model year “bins” to have too few samples to make meaningful comparisons. For example, the model 1981 through 1984 LDGT2s only had 4 vehicles, none of which were high emitters. With small sample sizes, a single test with conflicting outcomes (e.g. fail MA31 but pass IM240) can have a large effect on the failure rates and excess emissions for a particular “bin”. For this reason, the analyses focused more on comparing overall values for the entire dataset instead of individual “bins”.

10.2 Using the AZ 2% random dataset to determine MA31 test effectiveness

The most direct method of determining the effectiveness of the MA31 test in the Massachusetts I&M program would be to perform back-to-back MA31 and IM240 tests on vehicles from the Massachusetts fleet. However, because there wasn’t an existing IM240 test lane available in Massachusetts for testing, this was not a practical option.

One alternative was to use the AZ study dataset, which contains back–to-back MA31 and IM240 results for 612 vehicles from the Arizona fleet, for the analysis. However, since this dataset was designed to contain an equal distribution of low, medium, and high emitting vehicles (for the conversion factor analysis), it was not appropriate for determining the effectiveness of the MA31 test. In a typical vehicle fleet, the large majority of vehicles are low emitters that don't require repairs and generally don't fail the emissions test (i.e. they don't have excess emissions). The majority of excess emissions come from a small number of higher emitting vehicles whose emissions readings and test results will have a large bearing on the effectiveness of the emissions test. Although not as many high emitting vehicles were tested as anticipated, the AZ study dataset was not intended to represent a typical vehicle population and, therefore, should not be used directly to determine the test effectiveness. A dataset containing a larger proportion of high emitters would likely overestimate the potential effectiveness of the test.

Another alternative was to take a cross-section of MA31 test results from the Massachusetts I&M program and predict IM240 scores from these data using regressions developed from the AZ study dataset and a Monte Carlo simulation. However, Sierra had concerns about using MA31 results from the Massachusetts program because actual emission levels and failure rates appeared to be lower than expected for the program. With these data biased towards lower emissions, the analysis would likely underestimate the potential effectiveness of the MA31 test. DEP is investigating possible sources for the lower than expected emission scores with the Massachusetts program data. Likely reasons for this are quality assurance/quality control issues such as in-use MASS99 equipment problems, improper test delivery by inspectors, and motorist compliance (where the dirtier vehicles may not be showing up for inspection). These issues are outside the scope of this test effectiveness study and are being pursued separately by DEP.

The best available alternative for determining the MA31 test effectiveness for the Massachusetts fleet was to use an existing dataset containing a random sampling of 3,744 vehicles from Arizona's own I&M program evaluation (the AZ 2% sample). IM240 tests were performed for these vehicles under controlled conditions to assure proper calibration and performance of the test equipment and proper test procedure by the inspector. Using data from the 612-vehicle AZ study dataset, regressions were developed to predict MA31 scores from the IM240 scores in the AZ 2% dataset. A Monte Carlo simulation was then used to adjust these predicted scores to account for random variability observed between MA31 and IM240 scores in the original AZ study dataset.

There are limitations with using the AZ 2% sample dataset for the MA31 test effectiveness analysis. First of all, the Arizona fleet had already gone through multiple I&M cycles prior to their evaluation so it is reasonable to expect that their vehicles were better maintained than the Massachusetts fleet at this point. Secondly, the distribution of vehicles (model years, types, etc.) between Arizona and Massachusetts are likely different. Due to the difference in climate, the Massachusetts fleet tends to contain fewer older vehicles because many have been retired due to excess rust and/or poor cold weather performance. Finally, the Arizona 2% random IM240 dataset is now several years old. Consequently, the Massachusetts fleet may have a larger population of newer and cleaner technology vehicles than were represented in this dataset. The exact influence of these factors on the emissions results is difficult to predict and may have effects that cancel each other out. Nonetheless, this dataset provided the best opportunity to evaluate potential effectiveness of the MA31 test.

11.0 FUTURE WORK

This study tested 612 vehicles with MA31 and IM240 drive traces on MASS99 and IM240 test equipment. These data were used to determine MA31 to IM240 conversion factors for the Massachusetts program and the potential effectiveness of the MA31 test. To verify the accuracy of these results and conclusions and examine the effects of known limitations with this study, the following additional work is recommended:

- Options to Improve Test Effectiveness – This report presents three options that could be implemented to increase the test effectiveness of the program. Each of these options has potential disadvantages that need to be investigated before choosing and implementing the best one.

- Perform I&M Program Evaluation – This report investigated the effectiveness of the MA31 test by testing vehicles in a controlled environment. In addition to test effectiveness, other factors influence an I&M program's overall effectiveness, including proper calibration and test procedures by station personnel, and enforcing inspection requirements for vehicles not receiving timely tests. DEP plans to perform an evaluation of the I&M program to assess its overall effectiveness.

- Y-Intercept for MASS99 Conversion Factor Equation – The current MASS99 software uses a linear equation in the form of "y = mx" to convert raw MA31 scores to equivalent IM240 scores. This form of linear equation does not allow for a y-intercept and is known as "forcing the regression through the zero point". The regressions performed in this report on the final AZ study dataset were all forced through the zero point to accommodate the limitations of the existing conversion factor equation. A regression equation that includes a y-intercept (y = mx + b) is normally the more accurate form for a simple linear regression. The efficacy (in terms of increasing conversion accuracy) and costy of adding a y-intercept to the conversion factor equation in the MASS99 software should be investigated.

- Massachusetts IM240 Test Lane – In 2003, DEP will be installing an IM240 test lane at MassBay Community Technical College using equipment donated by Rhode Island's Department of Environmental Management. The SPX MASS99 system used in the AZ study test lane will also be set up. This facility will allow for side-by-side IM240 and MA31 tests so that results from the study can be verified.

APPENDIX A:
IM Program Cutpoints

Vehicle Type	Model Year Groups	EPA IM240 Start-Up *			EPA IM240 Final		
		HC	CO	NOx	HC	CO	NOx
LDGV	84-90	2.0	30	3.0	0.8	15	2.0
	91-95	1.2	20	2.5	0.8	15	2.0
	96+	0.8	15	2.0	0.6	10	1.5
LDGT1	84-87	3.2	80	7.0	1.6	40	4.5
	88-90	3.2	80	3.5	1.6	40	2.5
	91-95	2.4	60	3.0	1.6	40	2.5
	96+ (>3750 LVW)	1.0	20	2.5	0.8	13	1.8
	96+ (<3750 LVW)	0.8	15	2.0	0.6	10	1.5
LDGT2	84-87	3.2	80	7.0	1.6	40	4.5
	88-90	3.2	80	5.0	1.6	40	3.5
	91-95	2.4	60	4.5	1.6	40	3.5
	96+ (>3750 LVW)	2.4	60	4.0	0.8	15	2.0
	96+ (<3750 LVW)	1.0	20	2.5	0.8	13	1.8

- The Massachusetts MA31 cutpoints points = EPA IM240 Start-Up cutpoints

APPENDIX B:
Monte Carlo Simulation Report

RWCrawford ***Energy Systems / Analysis & Economics***

Robert W. Crawford ♦ 2853 S. Quail Trail ♦ Tucson, AZ 85730-5627 ♦ 520/546-1490
email ◊ rinconranch@earthlink.net
fax ◊ 520/546-1513

7 September 2001

Massachusetts Correlation Study:
Documentation of IM240/MA31 Emission Correlations

This brief report documents the results of an analysis conducted to determine the relationship between emission scores on the IM240 short test and those on the MA31 test. The statistical relationships are termed emission "correlations" in this report, as a matter of shorthand, although they are stated in the form of regression equations, and not correlation coefficients.

The statistical relationships, or emission correlations, can be used to estimate MA31 scores from IM240 scores. A primary application of the correlations will be in a Monte Carlo simulation of MA31 test data, in which the distribution of MA31 scores will be estimated as a function of the IM240 test result. Therefore, a thorough residuals[1] analysis was conducted, both to validate the usual assumptions of regression analysis and to develop recommendations for the modeling of residuals.

Methodology

The emissions data used here originates from a testing program conducted in Arizona, in which a sample of passenger cars and light trucks has been subjected to IM240 and MA31 tests. Higher emission vehicles are intentionally over-represented in the data in a effort to improve the density of information at higher emission levels. The dataset consists of a total of 612 vehicles, including 326 passenger cars and 286 light duty trucks. The analysis was conducted separately for cars and light trucks.

Emissions data typically display a right-skewed distribution of values, as is commonly encountered whenever the physical measurement must be a non-negative value. The

[1] Residuals are defined in this study as the predicted value minus the observed value; positive residuals indicate that the emissions correlation over-predicts actual emissions.

distributions for IM240 and MA31 emissions are approximately log-normal – i.e., normally distributed when transformed to logarithms – although the data exhibit some degree of asymmetry and tend to over-populate the tails of the distribution compared to what is expected in the log-normal distribution. These features are common in real data and mean, merely, that any statistical distribution will be an approximation to reality.

Regression analysis is based on the assumption that the residuals to the regression are normally distributed and independent of the explanatory variables. If this assumption is valid, then the coefficients estimated by ordinary least squares (OLS) will be unbiased with respect to the population (or "true") values. If the residuals are significantly skewed compared to the normal distribution, the coefficient estimates will deviate from the population values, in the same conceptual sense that the mean value of right-skewed data will be greater than the median (or most-likely) value. If the residuals are not independent of the explanatory variables – i.e., if they increase or decrease in average size as one moves across the range of one or more explanatory variable – then there is some degree of mis-fit in the model. If the dispersion varies systematically, the coefficient estimates will be distorted as the regression gives disproportional weight to the range(s) of the data where the residuals tend to be the greatest.

It is always useful to test the residuals for normality and independence, although it is seldom possible to achieve complete conformance with the key assumptions made in regression analysis. These considerations are of increased importance in this study, because the Monte Carlo simulations in which the results will be used depend on both the expected value for MA31 scores (the regression predictions) and the random distribution of values (the residuals).

A log-log formulation (natural logarithms) was selected for the regression equations after considering several alternatives, and this formulation was found to fit the data relatively well in terms of both the overall trends in the data and the characteristics of the residuals. The general form of the regression model is the following:

$$\log(\mathrm{MA31}_i) = A + B*\log(\mathrm{IM240}_i) + C*\log(\mathrm{IM240}_j) \qquad (1)$$

where i is the pollutant in question and j is another pollutant. This equation is mathematically equivalent to:

$$\mathrm{MA31}_i = \exp(A) * (\mathrm{IM240}_i)^B * (\mathrm{IM240}_j)^C \qquad (2)$$

The B coefficients are all less than 1.00, with values ranging from 0.75 to 0.99. The overall shape of the emission correlation is that MA31 scores increase with IM240 scores, but at a slower rate as the IM240 score increases. The C coefficients range from 0.05 to 0.23 and introduce an adjustment to the overall trend.

As an example, the passenger car model for HC is a function of IM240 scores for HC and NO_x. The "same pollutant" term of the equation dominates the prediction of MA31 scores, while the "other pollutant" part provides a small, but statistically significant adjustment to the prediction that accounts for cross-correlations among the pollutants. Among the six different regressions models developed (HC, CO, and NO_x, separately for cars and light trucks), four of the models involve a second pollutant, while two of the models involve only the same pollutant term. Three pollutant models were considered, but in no case were all terms statistically significant.

The process of the analysis was the following:

- Estimate regression equations for each of the six groups, selecting the best one or two pollutant model having coefficients that are statistically significant at the 0.05 level. Data points with zero values for the pollutant terms involved were necessarily deleted from the sample used to estimate each log-log equation.

- Formulate a chi^2 test for normality of the residuals. The residuals were binned into 10 groups such that the bins would be equally populated if the residuals were normally distributed. The actual distribution of residuals was tested against this expected distribution using a conventional chi^2 test. The deviations from normality were, in all of the final models, found to be not so large that they were unlikely to occur by chance.

- The parameters of a normal distribution were fit to the residuals. The mean of this distribution must be zero as a mathematical requirement of the OLS procedure, while the standard deviation is non-zero.

- A test was conducted for outliers by identifying any data point whose residual was greater in absolute value than 4 times the residual standard deviation. A deviation greater than 4-sigma (plus or minus) will occur by chance in approximately 1 in 1500 cases, and points found to lie in this range are highly likely to represent erroneous data. Outliers were deleted and the regression equations re-estimated until no further outliers were identified.

- The residuals of the final models were then examined for evidence of a systematic trend with the same-pollutant IM240 score (the dominant explanatory variable). A smoothing process was applied to the residuals data to reduce the degree of random variation and to better reveal any systematic trend. The residuals were first sorted in increasing order of IM240 scores. An 11-point running average window was passed across the sorted data, with the calculated average associated with the IM240 score for the mid-point of the window. The smoothed residuals data were then plotted against IM240 and examined visually for evidence of any systematic trend. Some evidence of systematic variation was seen in 3 of the six regression groups. These instances are discussed further in the section on modeling the residual distributions.

Emission Correlation Results

The results of the emission correlation analysis are summarized in Table 1. Using the CO correlation results for passenger cars as an example, we see that a total of 315 data point had non-zero emission scores for the pollutants involved – in this case, MA31 CO and IM240 CO. Three data points were deleted as outliers, leaving 312 points for estimating the correlation equation. The equation involved CO 240 as the explanatory value, with a coefficient of 0.9363, and achieved an R^2 statistic of 0.699 in log-space. The chi^2 test for normality of the residuals indicated that the departures from strict normality could occur by random chance in 15.5 percent of samples. Therefore, we cannot reject the hypothesis that the residuals are normally distributed. The standard deviation of the residuals was 0.8928, as measured in log-space. The residuals data displayed a systematic deviation at low and high ends of the IM240 range.

Table 1: Summary of Emission Correlation Results

	Passenger Cars		Light Duty Trucks	
HC	**Sample Size**	322	**Sample Size**	282
	Outliers deleted	4	Outliers deleted	3
	N	318	N	279
	Regression		**Regression**	
	R^2	0.707	R^2	0.740
	Intercept (t-value)	-0.2571 (4.24)	Intercept (t-value)	-0.1440 (3.11)
	HC 240 (t-value)	0.8618 (27.6)	HC 240 (t-value)	0.8248 (28.1)
	Residuals		**Residuals**	
	Normality		Normality	
	(prob>chi^2)	.077	(prob>chi^2)	.1433
	Standard Deviation	0.7966	Standard Deviation	0.7169
	Systematic Trend	none	Systematic Trend	low
CO	**Sample Size**	315	**Sample Size**	275
	Outliers deleted	3	Outliers deleted	4
	N	312	N	271
	Regression		**Regression**	
	R^2	0.699	R^2	0.756
	Intercept (t-value)	-0.1770 (2.37)	Intercept (t-value)	-0.2684 (3.09)
	CO 240 (t-value)	0.9363 (26.8)	CO 240 (t-value)	0.9930 (28.9)
	Residuals		**Residuals**	
	Normality		Normality	
	(prob>chi^2)	0.155	(prob>chi^2)	0.149
	Standard Deviation	0.8928	Standard Deviation	0.7424
	Systematic Trend	low/high	Systematic Trend	low
NO_x	**Sample Size**	325	**Sample Size**	286
	Outliers deleted	5	Outliers deleted	6
	N	320	N	280
	Regression		**Regression**	
	R^2	0.777	R^2	0.759
	Intercept (t-value)	-0.1038 (3.93)	Intercept (t-value)	-0.1846 (5.67)
	NO_x 240 (t-value)	0.8929 (33.3)	NO_x 240 (t-value)	0.9331 (29.6)
	Residuals		**Residuals**	
	Normality		Normality	
	(prob>chi^2)	0.155	(prob>chi^2)	0.105
	Standard Deviation	0.4712	Standard Deviation	0.4698
	Systematic Trend	low	Systematic Trend	none

Notes: 1. Regression coefficients are for (natural) log-log model.
2. Residuals standard deviation is in unit of natural log.
3. Systematic trend for residuals is with respect to IM240 scores for the pollutant. The IM240 range(s) displaying apparent deviations from mean=zero are indicated by codes: low, mid, or high.

Modeling Residual Distributions

The apparent extent of systematic behavior found in the residuals is generally small to modest. To the extent it exists, the systematic behavior will lead to under- or over-prediction of MA31 emissions in some ranges of IM240 scores. In three cases (passenger car CO and NO_x and light truck CO), systematic behavior occurs at the low end of the IM240 range, where predicted MA31 scores are also small and the effect of biases will be greatly reduced. For example, the low range for passenger car CO (IM240 score < 1.4 gm/mi) exhibits a mean log residual of +0.07, which is equivalent to an average bias of +7 percent. This range has predicted MA31 scores of typically 0.5 gm/mi or less; the corresponding observed MA31 would tend to be smaller still. Effects of this direction and size will probably have little impact on studies of short-test failure rates.

Systematic behavior also exists at the high end of the IM240 range for passenger car CO. Here, for IM240 scores in excess of 19 gm/mi, the emissions correlations can under-predict MA31 CO by as much as 20 to 30 percent. Because observed MA31 emissions would be higher than predicted, these biases also will have little or no effect on studies of short-test failure rates.

In general, the presence of systematic residuals in some instances is likely to have little or no effect on the intended uses for these correlations. It should be appropriate for many purposes to model the residuals on a pooled basis as a normal distribution with mean zero and standard deviation as shown in the table.

Greater precautions against the possibility of systematic behavior can be taken by modeling the residuals separately in three ranges (low, middle, and high) for each group, as shown in Table 2. Sub-populations have been defined as follows:

- Low range = lowest 1/7th of the IM240 scores
- Middle range = middle 5/7ths of the IM240 scores
- High range = highest 1/7th of the IM240 scores

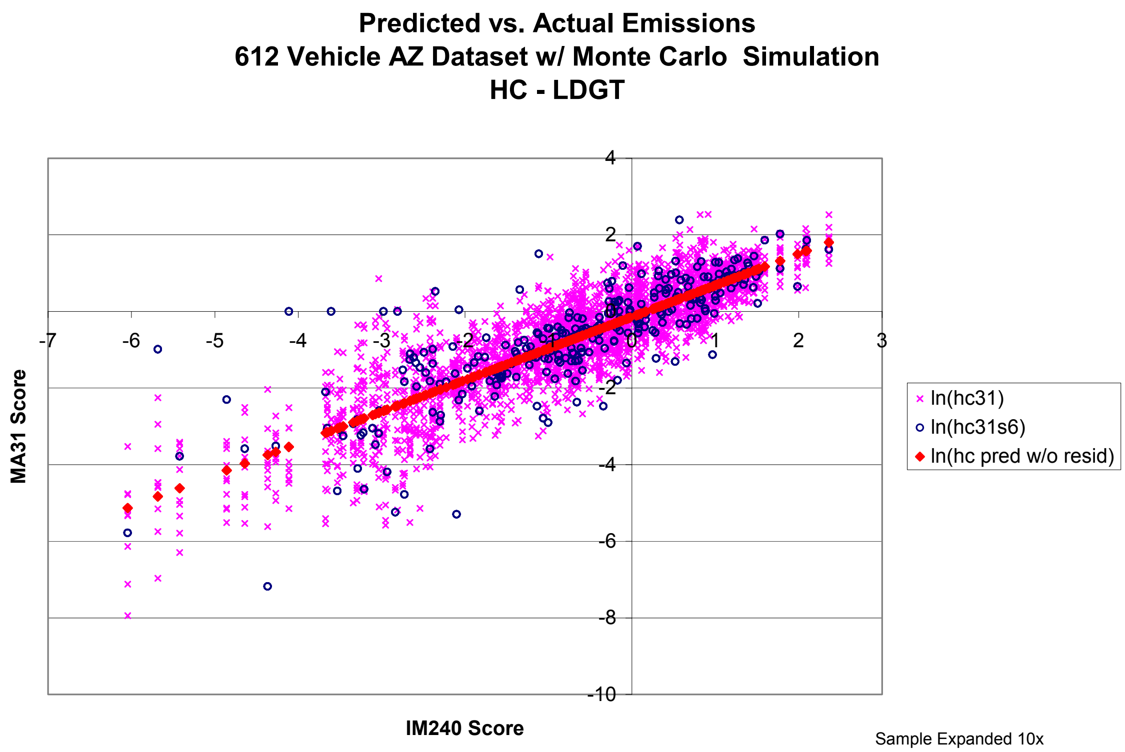
These generic criteria are based on the three cases where systematic behavior in residuals was observed (passenger car CO and NO_x and light truck CO). The means and standard deviations (log-space) are then calculated for the residuals in each sub-population and tabulated in the table.

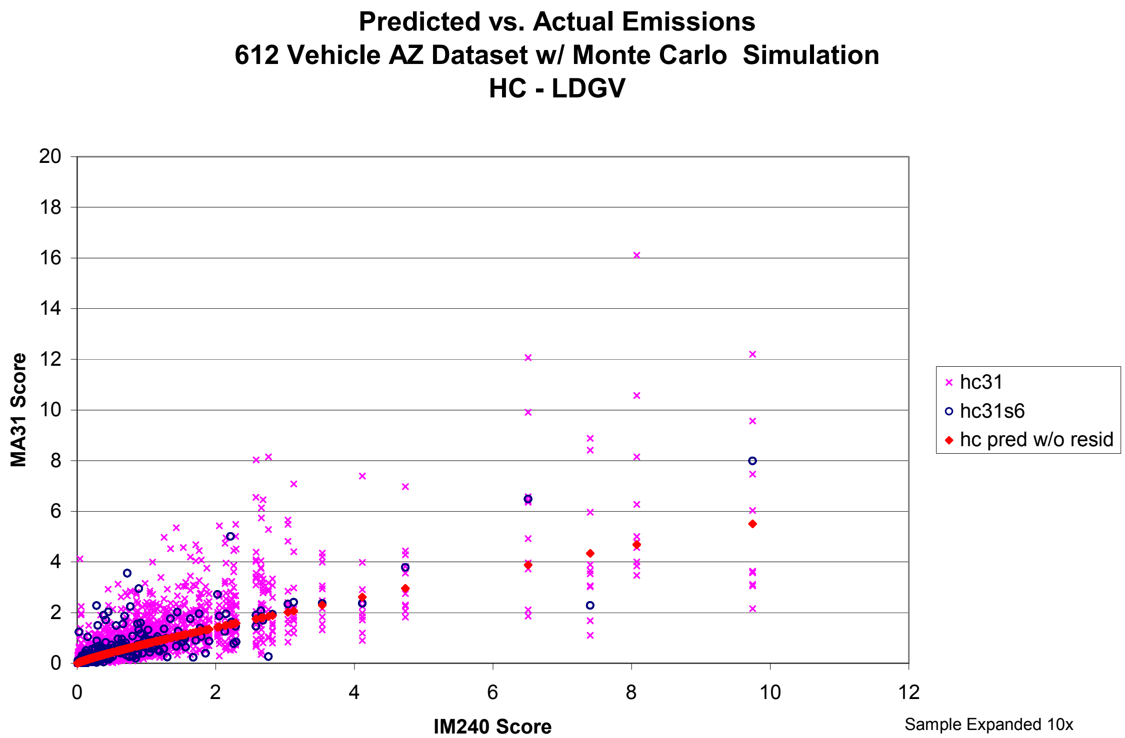
Using passenger car CO as an example, cars with IM240 CO below 1.370 gm/mi would be identified as the low-range sub-population and assigned residuals drawn from the population N(0.0718, 1.1595). Cars with IM240 CO above 19.12 gm/mi would be identified as the high-range sub-population and assigned residuals from the population N(0.0250, 0.7613). One can see that there is a tendency for the residuals standard deviation to decrease as one goes from the low to high sub-populations. This suggests the accuracy with which the distribution of emission values can be simulated will be improved by dividing the data into the three sub-populations. This may have little effect, however, on the overall results, if the most important range of the data is in the middle.

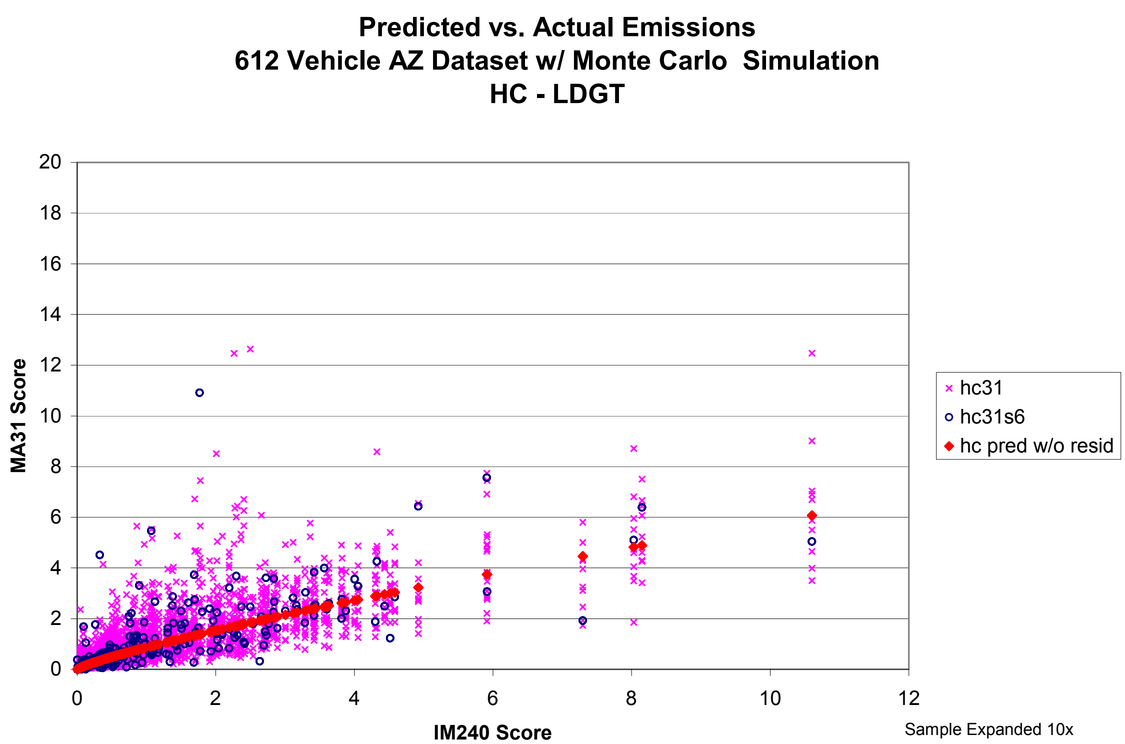
Table 2: Parameters for Modeling Residual Distributions		

	Passenger Cars			Light Duty Trucks		
	gm/mile	Mean Log	StdDev Log	gm/mile	Mean Log	StdDev Log
HC	**All**	**0.0000**	**0.7966**	**All**	**0.0000**	**0.7169**
	< 0.060	0.1356	1.2377	< 0.097	-0.0024	1.1334
	Middle	-0.0482	0.7163	Middle	0.0080	0.6636
	> 1.255	0.1085	0.5913	> 2.680	-0.0152	0.3739
CO	**All**	**0.0000**	**0.8928**	**All**	**0.0000**	**0.7424**
	< 1.370	0.0610	1.1887	< 2.457	0.1932	1.1877
	Middle	-0.0143	0.8468	Middle	-0.0779	0.6381
	> 19.12	0.0096	0.7835	> 31.87	0.1922	0.5836
NOx	**All**	**0.0000**	**0.4712**	**All**	**0.0000**	**0.4698**
	< 0.406	-0.0084	0.7996	< 0.670	0.0329	0.7157
	Middle	0.0062	0.3989	Middle	-0.0040	0.4183
	> 2.657	-0.0224	0.3621	> 3.915	-0.0131	0.4159

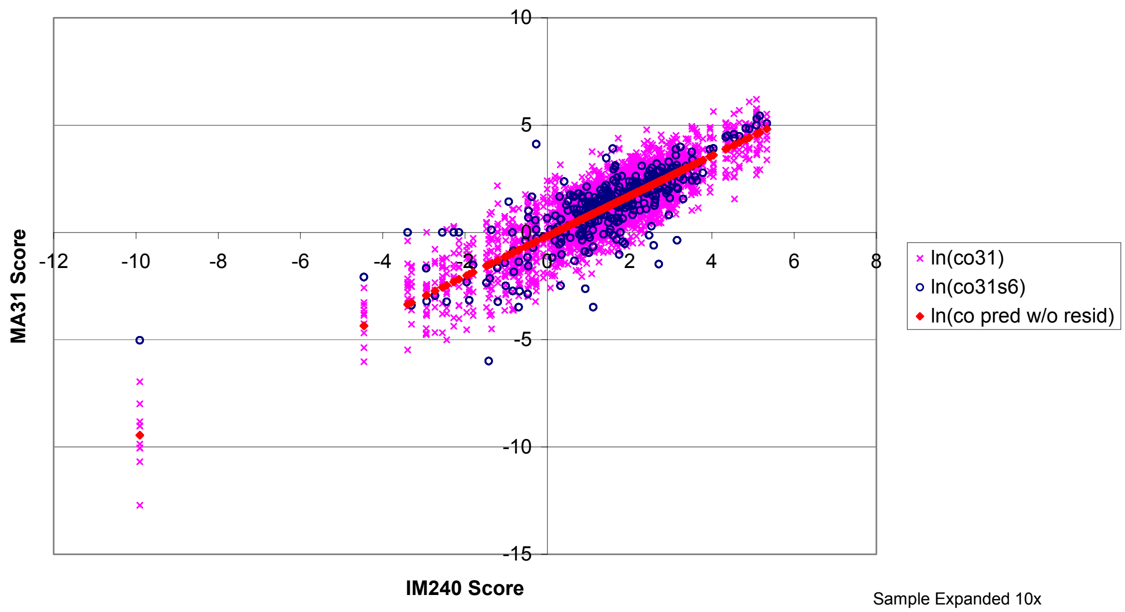
APPENDIX C:
Monte Carlo Simulation for predicting MA31 scores from IM240 scores



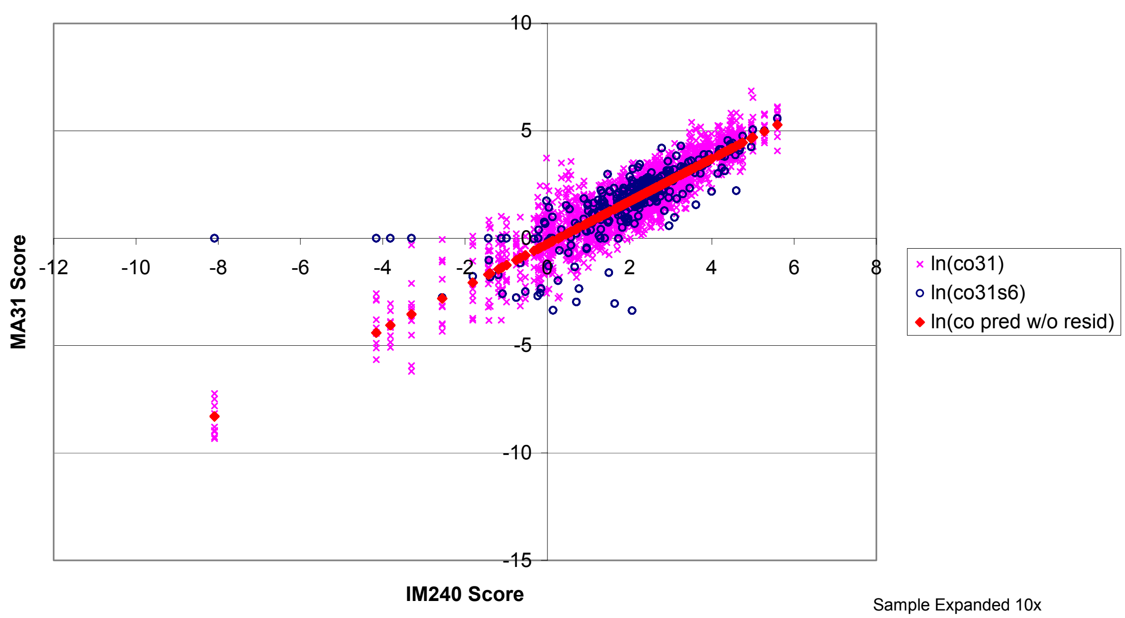

Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
HC - LDGV
MA31 Score
IM240 Score
hc31
hc31s6
hc pred w/o resid
Sample Expanded 10x

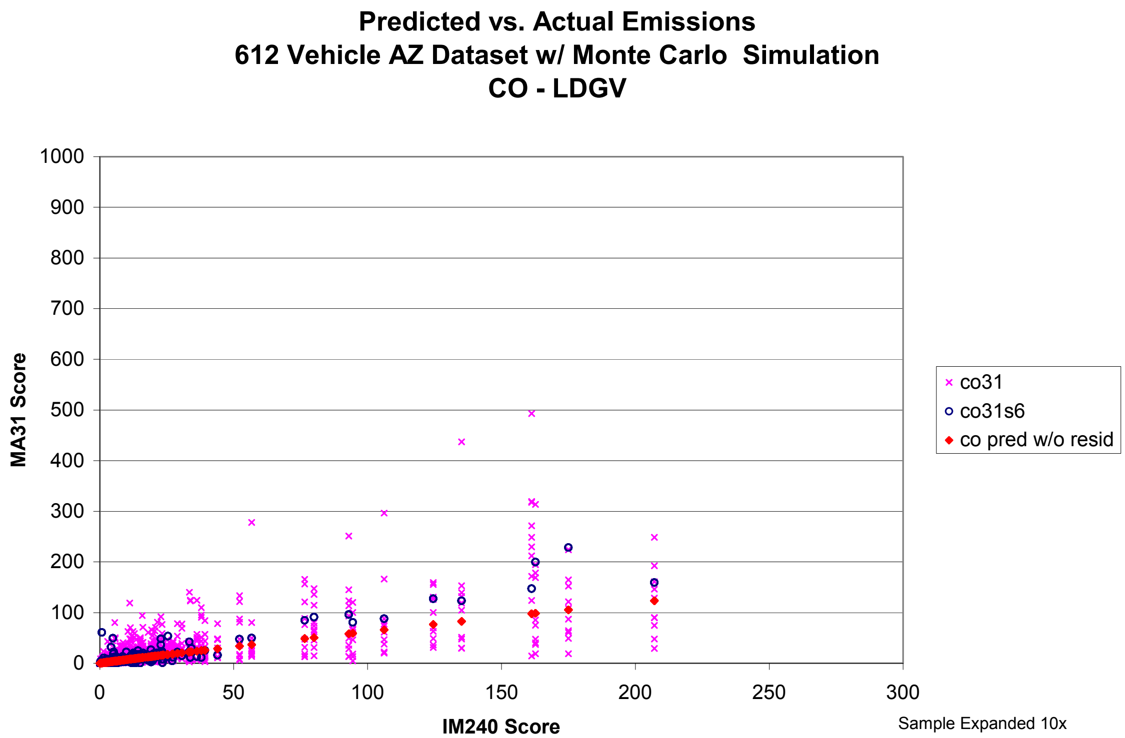

Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
HC - LDGT
MA31 Score
IM240 Score
hc31
hc31s6
hc pred w/o resid
Sample Expanded 10x

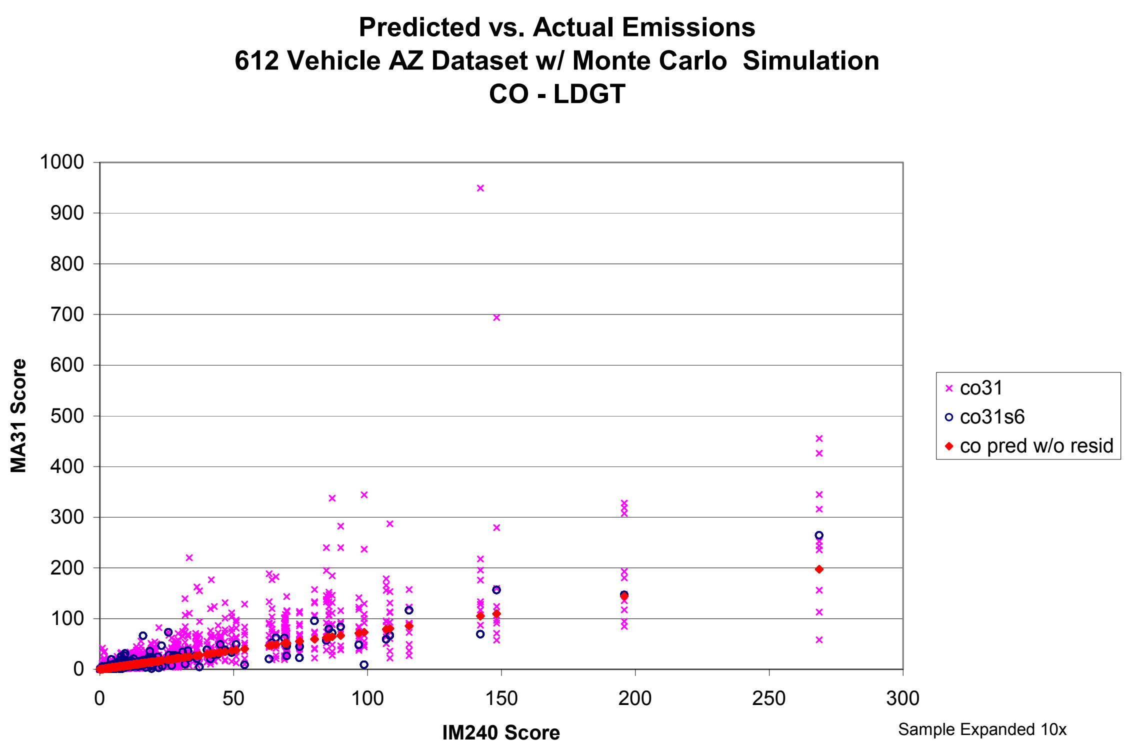
Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
CO - LDGV



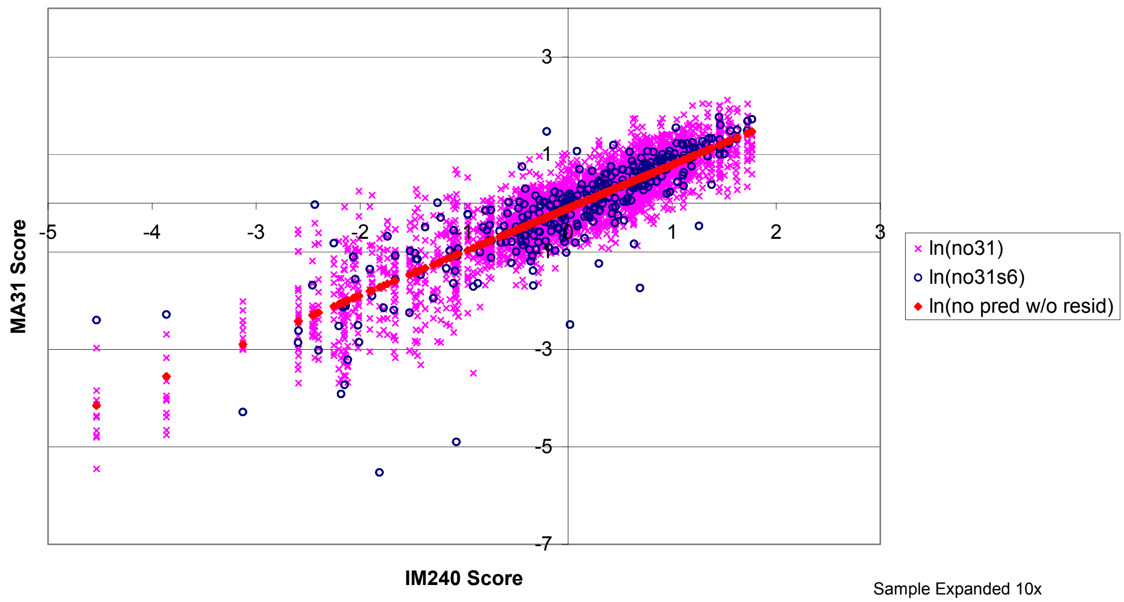
Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
CO - LDGT



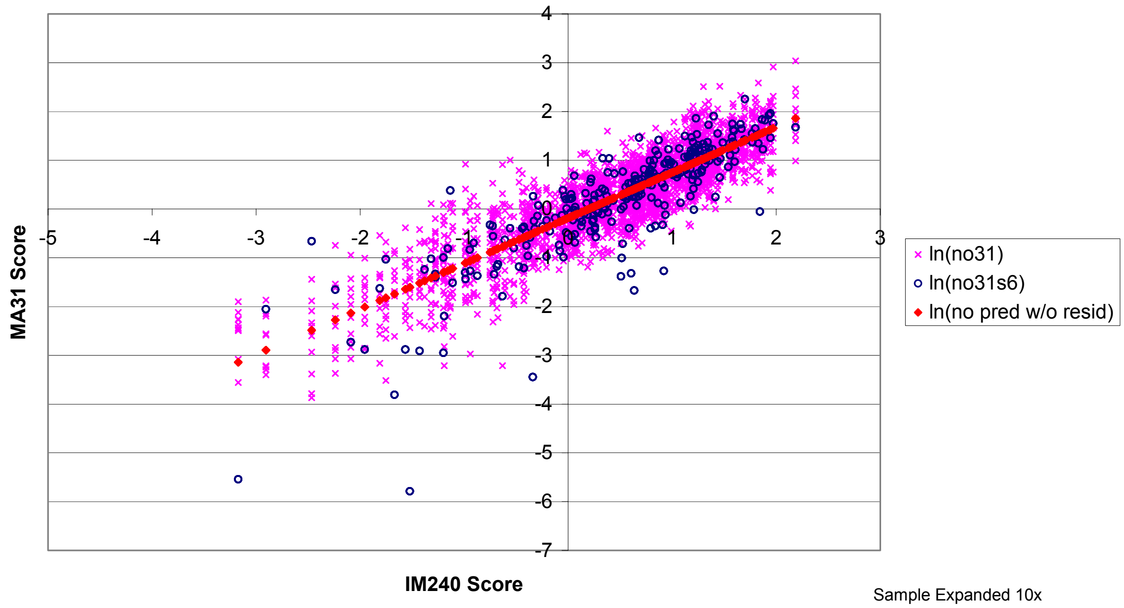

Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
CO - LDGV
MA31 Score
IM240 Score
co31
co31s6
co pred w/o resid
Sample Expanded 10x
0
100
200
300
400
500
600
700
800
900
1000
0
50
100
150
200
250
300


Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
CO - LDGT
MA31 Score
IM240 Score
co31
co31s6
co pred w/o resid
Sample Expanded 10x
0
100
200
300
400
500
600
700
800
900
1000
0
50
100
150
200
250
300

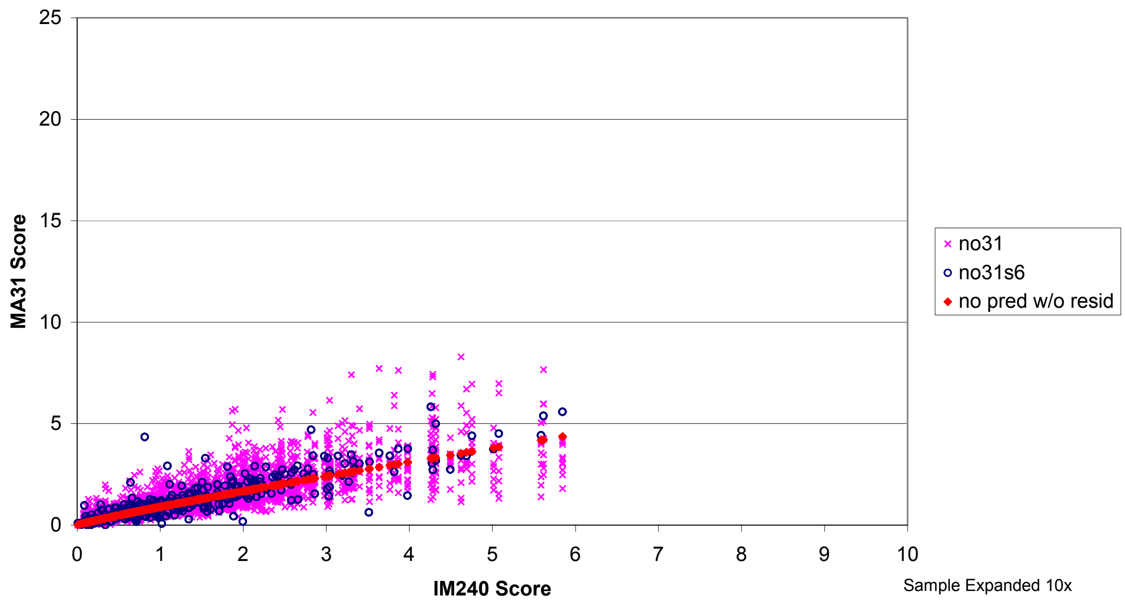
Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
NO - LDGV



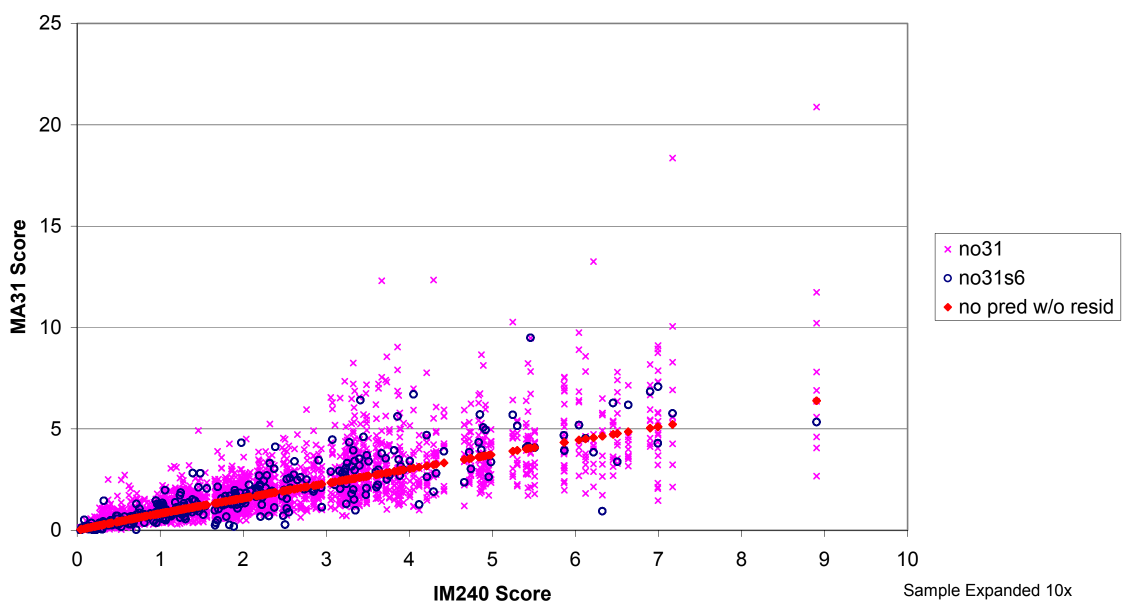
Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
NO - LDGT



Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
NO - LDGV



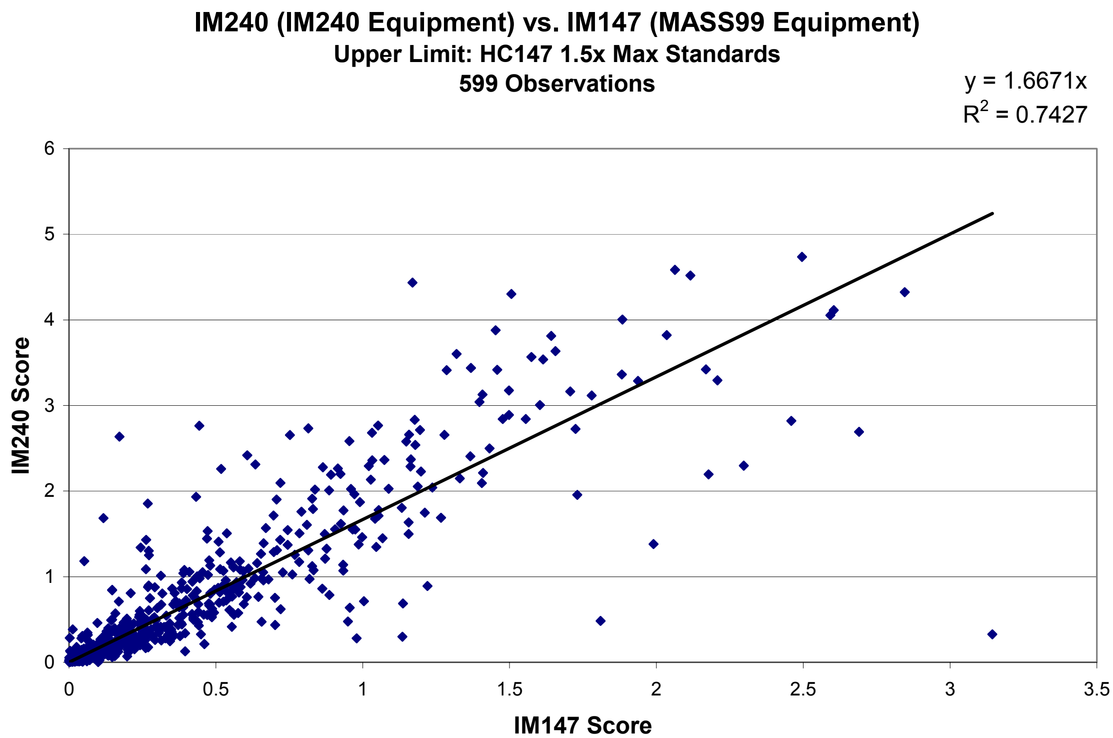
Predicted vs. Actual Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
NO - LDGT

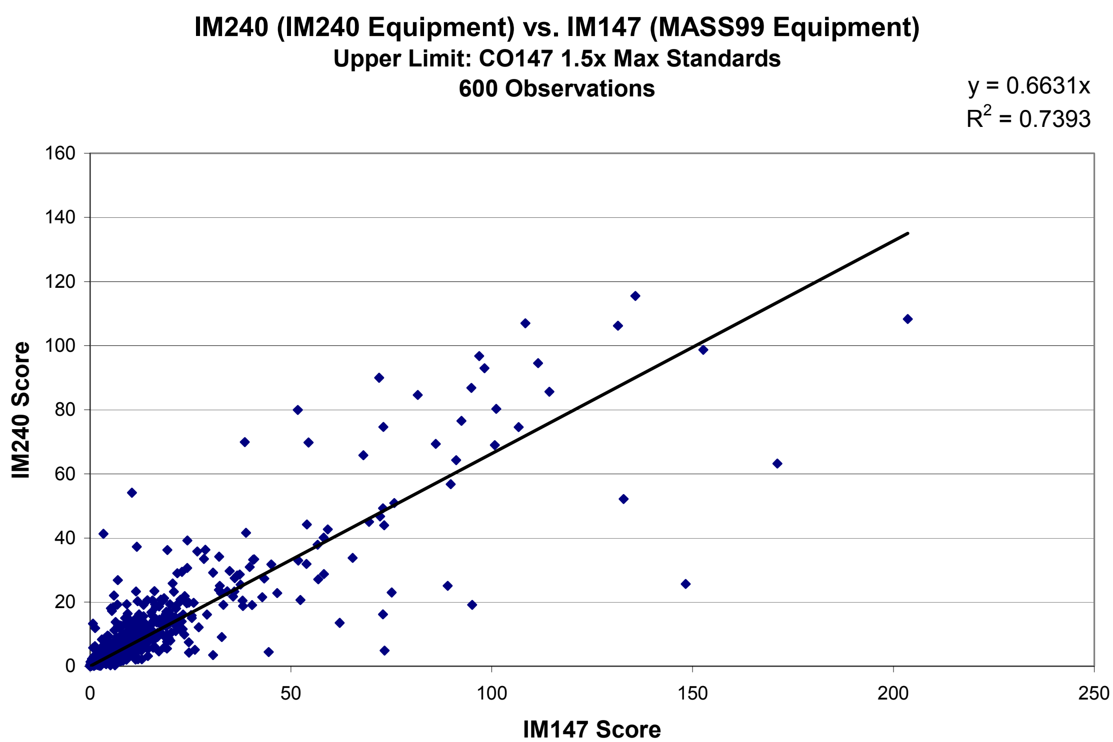


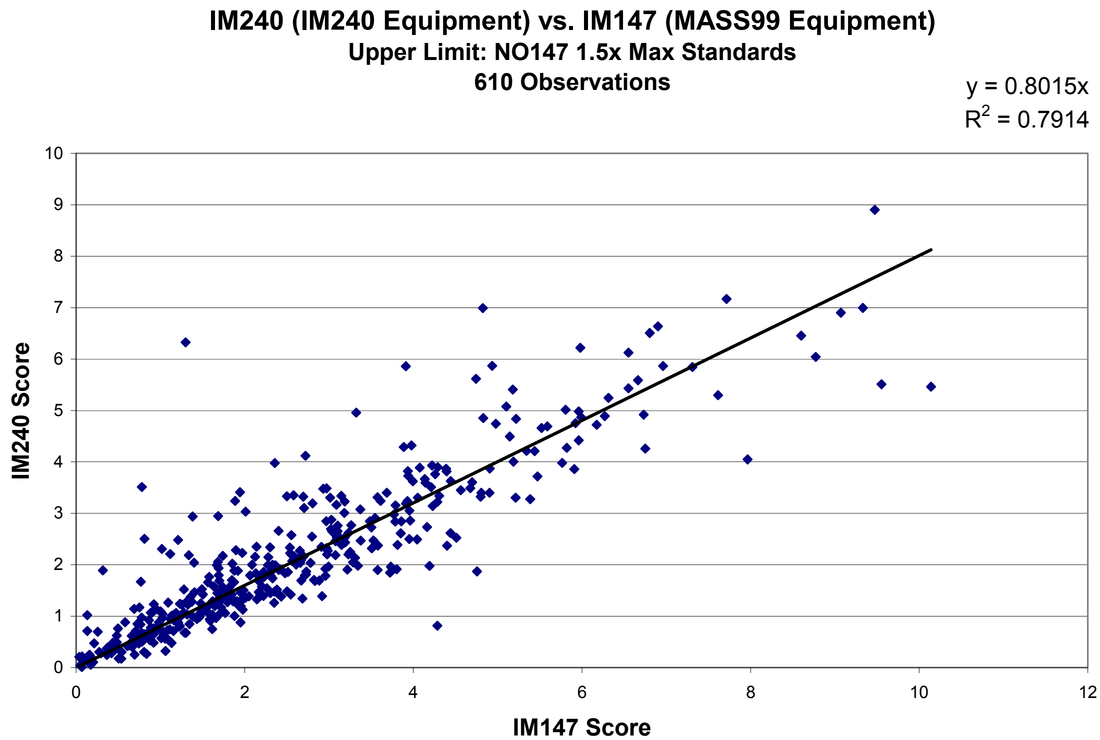
APPENDIX D:
Mileage Accumulation Rates from EPA MOBILE 6.2 Users Guide

Vehicle Age (years)	**Mileage Accumulation Rates** (miles per year)	
	Passenger Cars (LDGV)	**Trucks** (LDGT1, LDGT2)
1	14,910	19,496
2	14,174	18,384
3	13,475	17,308
4	12,810	16,267
5	12,178	15,260
6	11,577	14,289
7	11,006	13,352
8	10,463	12,451
9	9,947	11,584
10	9,456	10,752
11	8,989	9,955
12	8,546	9,194
13	8,124	8,467
14	7,723	7,775
15	7,342	7,118
16	6,980	6,496
17	6,636	5,909
18	6,308	5,356
19	5,997	4,839
20	5,701	4,357
21	5,420	3,909
22	5,152	3,497
23	4,898	3,120
24	4,656	2,777
25	4,427	2,470

APPENDIX E:
MA147 to IM240 Regressions

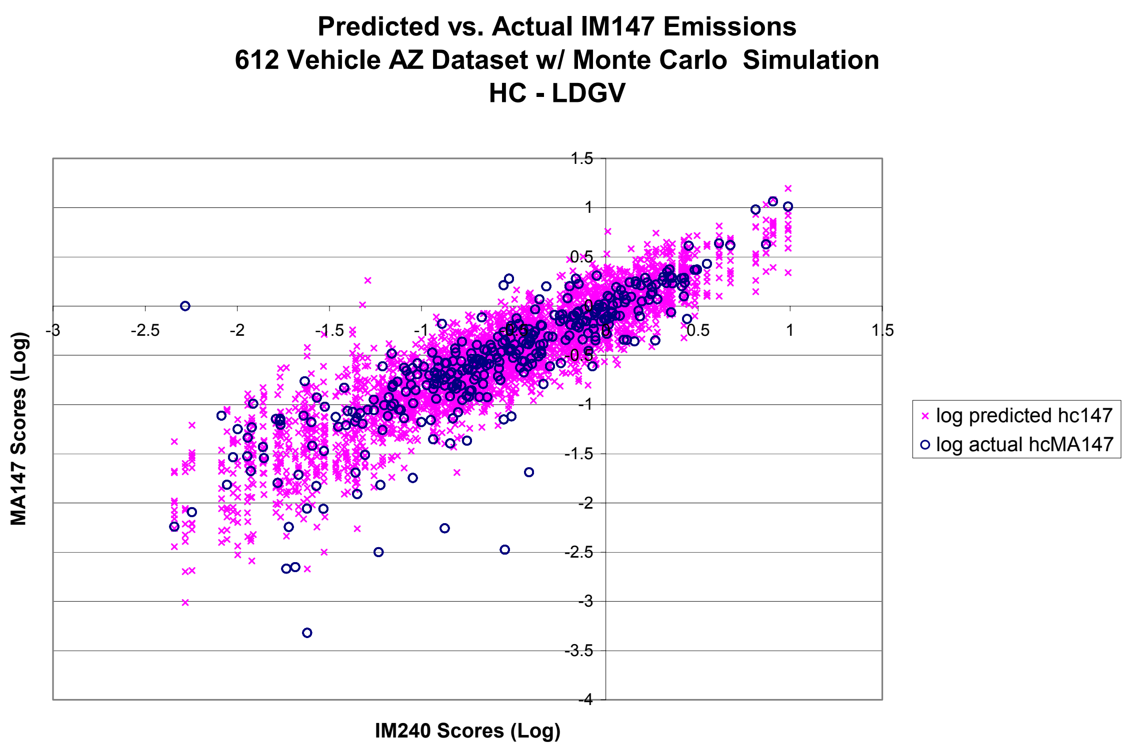


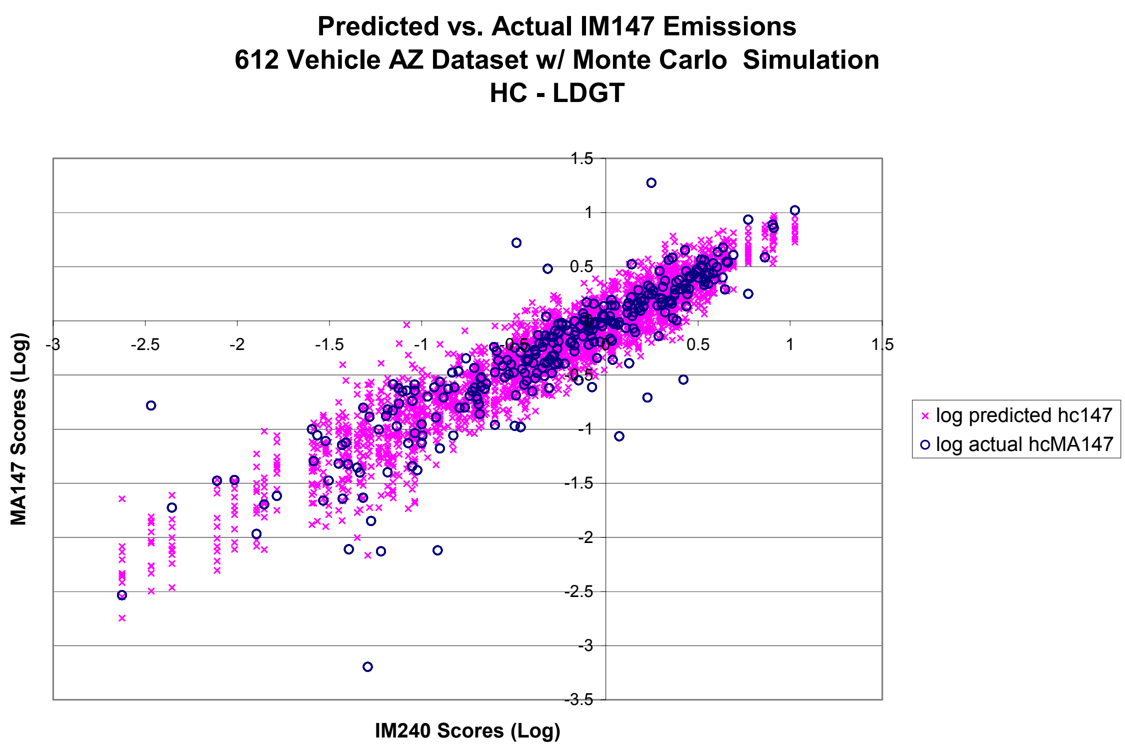


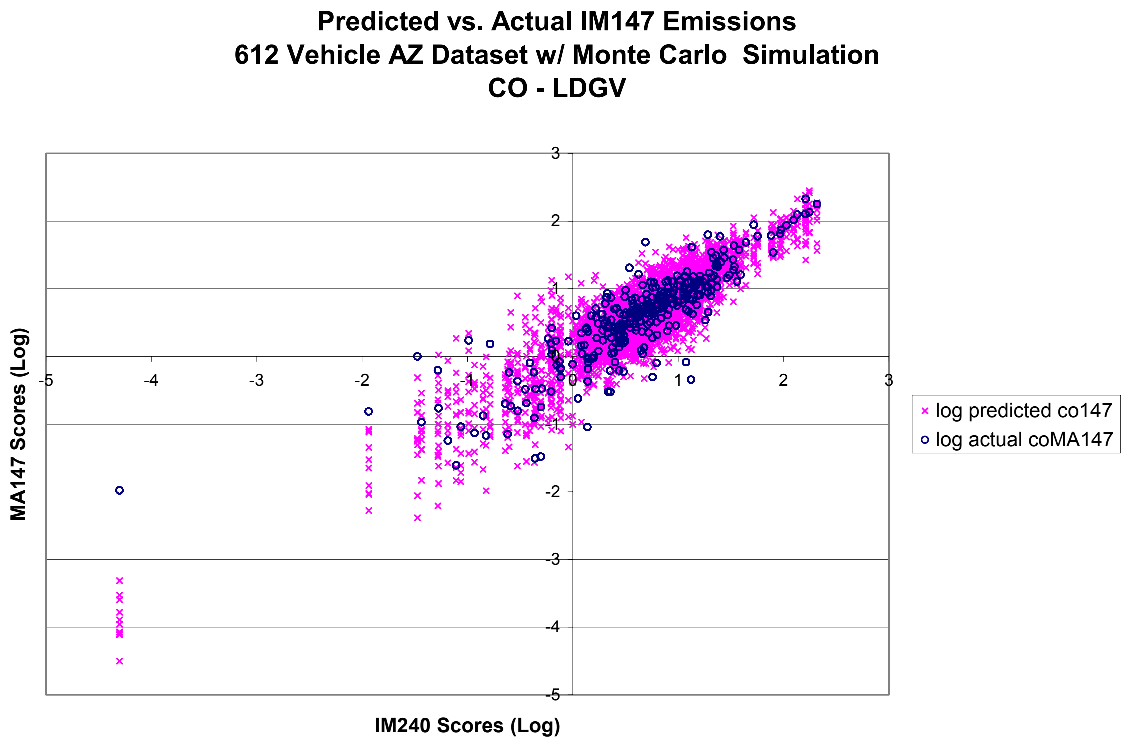

IM240 (IM240 Equipment) vs. IM147 (MASS99 Equipment)
Upper Limit: NO147 1.5x Max Standards
610 Observations
y = 0.8015x
R2 = 0.7914
IM240 Score
IM147 Score
0
1
2
3
4
5
6
7
8
9
10
0
2
4
6
8
10
12

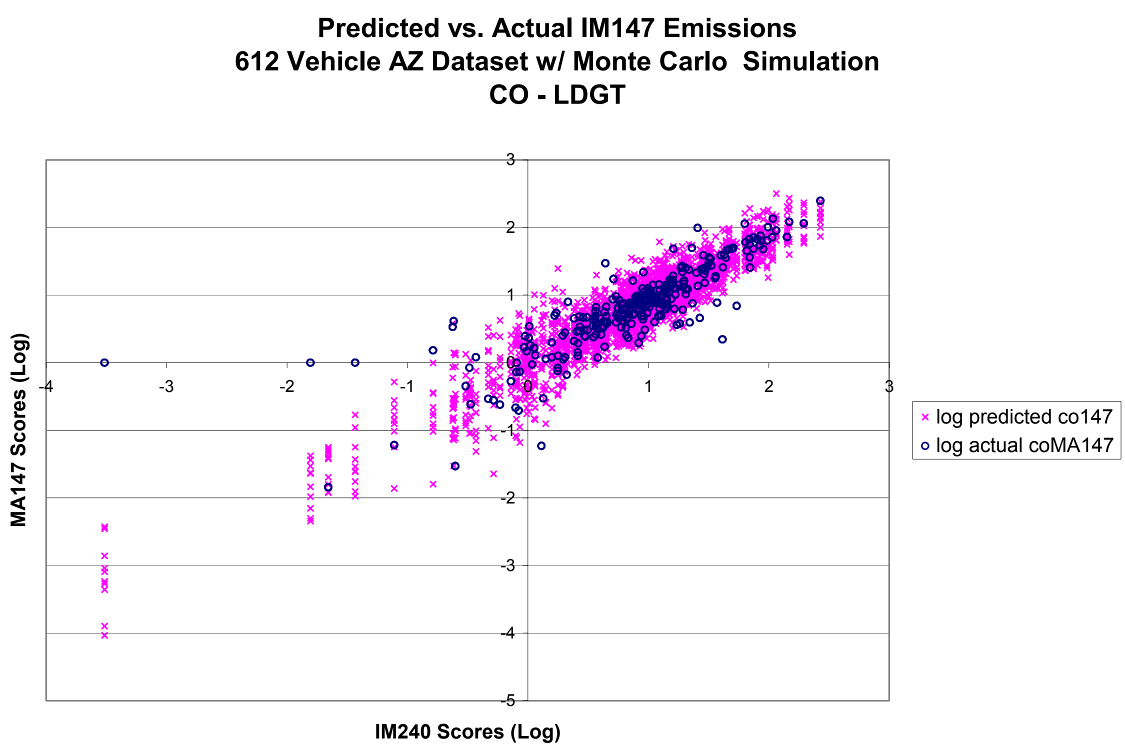
APPENDIX F:

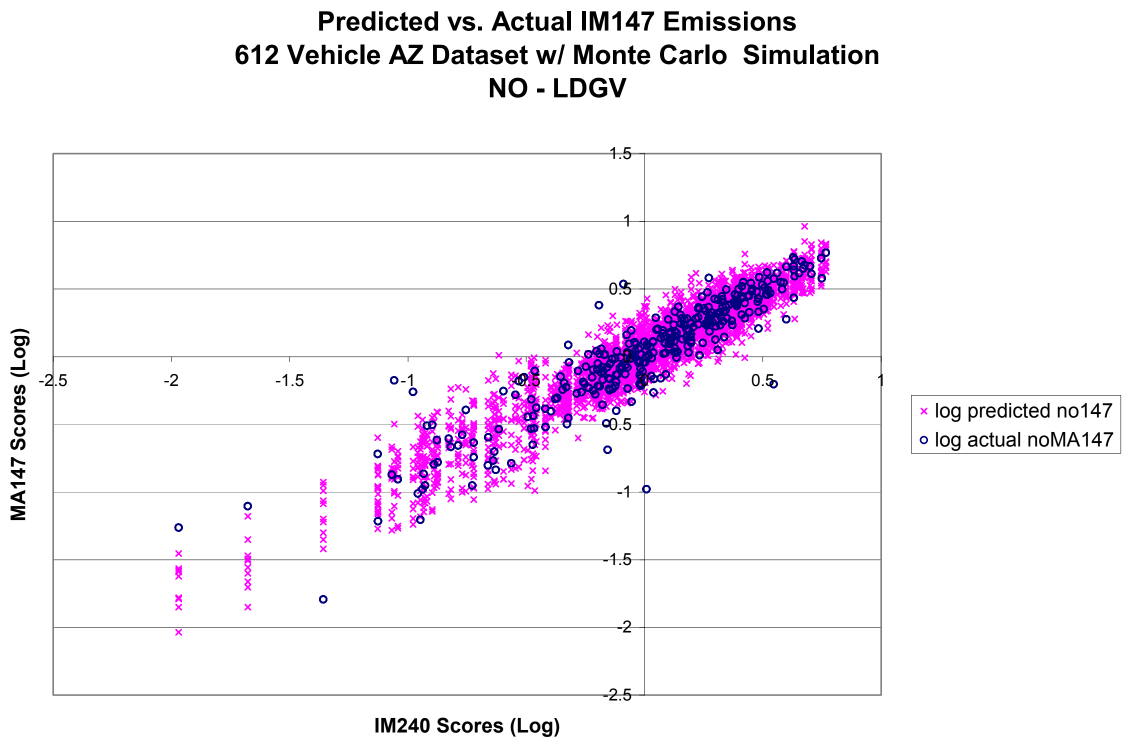
Monte Carlo Simulation for predicting MA147 scores from IM240 scores

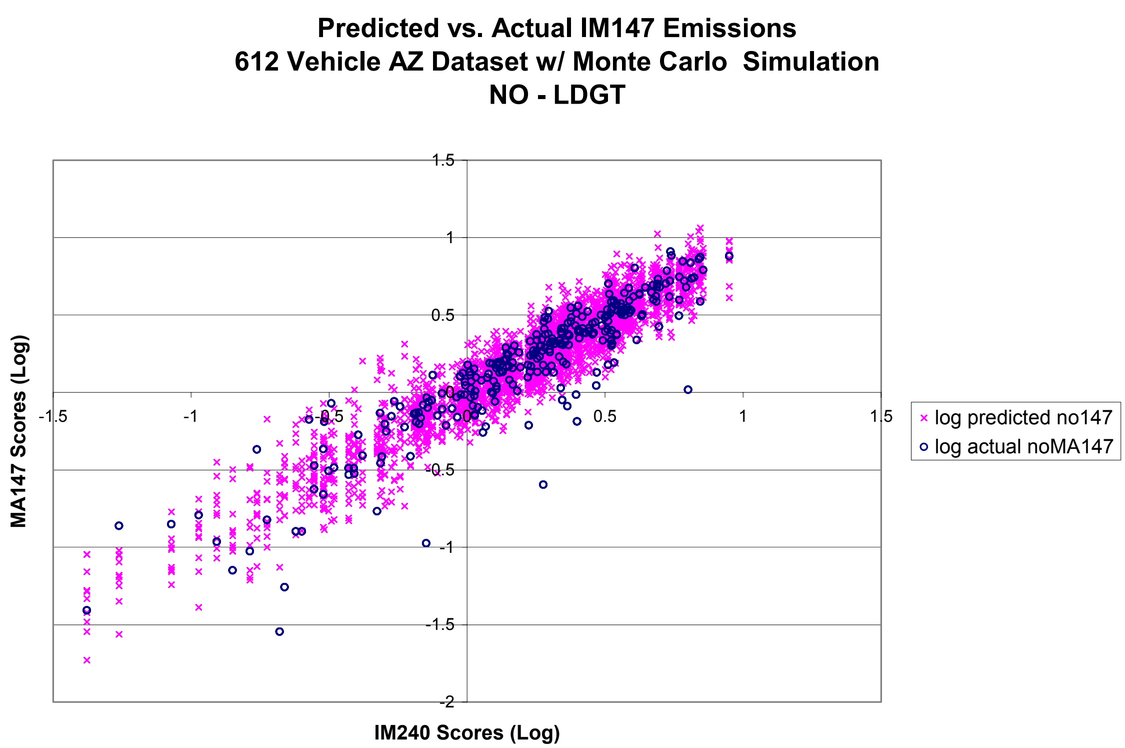





Predicted vs. Actual IM147 Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
CO - LDGV
MA147 Scores (Log)
IM240 Scores (Log)
log predicted co147
log actual coMA147
-5
-4
-3
-2
-1
0
1
2
3


Predicted vs. Actual IM147 Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
CO - LDGT
MA147 Scores (Log)
IM240 Scores (Log)
log predicted co147
log actual coMA147
-5
-4
-3
-2
-1
0
1
2
3


Predicted vs. Actual IM147 Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
NO - LDGV
MA147 Scores (Log)
IM240 Scores (Log)
log predicted no147
log actual noMA147


Predicted vs. Actual IM147 Emissions
612 Vehicle AZ Dataset w/ Monte Carlo Simulation
NO - LDGT
MA147 Scores (Log)
IM240 Scores (Log)
log predicted no147
log actual noMA147